DOCUMENT RESUME

ED 281 919 UD 025 468

AUTHOR Kennedy, Mary M.; And Others
TITLE The Effectiveness of Chapter 1 Services. Second Interim Report from the National Assessment of Chapter 1.
INSTITUTION Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO OR-87-501
PUB DATE Jul 86
NOTE 175p.
AVAILABLE FROM Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS *Compensatory Education; Disadvantaged Youth; Educationally Disadvantaged; Elementary Secondary Education; *Federal Programs; Individualized Education Programs; Program Effectiveness; Remedial Programs; *Special Education; *Special Programs
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1; *Elementary Secondary Education Act Title I

ABSTRACT

The following findings are presented in this second evaluation of Chapter 1 of the Education Consolidation and Improvement Act of 1981: (1) the achievement of disadvantaged students has improved since 1965, relative to the general population; (2) Chapter 1 students experience greater increases in achievement test scores than others; (3) students in Chapter 1 mathematics programs gain more than those in Chapter 1 reading programs; (4) students in early-elementary Chapter 1 programs gain more than those in later-grade programs; (5) evidence of program effects on student attitudes toward school is inconclusive; (6) the relationship between program costs and effects on achievement scores is unclear; (7) the achievement gap between disadvantaged and advantaged students widens during summer; (8) Title I summer programs, not academically rigorous, have not narrowed this gap; (9) students who discontinue Title I lose gains they made when receiving services; (10) Chapter 1 students with very low achievement scores maintain their relative academic positions but do not advance; (11) long-term effects of Chapter 1 programs on graduation rates, future education, or adult literacy are unknown; (12) attempts to identify particular project characteristics that improve student achievement test scores have been ineffective; (13) instructional practices likely to be effective in Chapter 1 programs have been recently identified. Appendices contain the statute requiring the assessment; an administrative status report; studies, data sources, and measures used in the report; achievement expressed in normal curve equivalents; and sample sizes for figures and tables. (Author/LHW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

4-16-87

THE EFFECTIVENESS OF CHAPTER 1 SERVICES

ED281919

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

OERI
Office of Educational Research and Improvement
U.S. Department of Education
Office of Research

NATIONAL ASSESSMENT OF CHAPTER 1 • 1986

UD 025 468

THE EFFECTIVENESS OF CHAPTER 1 SERVICES

by

Mary M. Kennedy
Michigan State University

Beatrice F. Birman
Randy E. Demaline
U.S. Department of Education

with the assistance of

Mary T. Moore
David E. Myers
Stephen Chaikind
William Strang
Adrienne von Glatz
DRC

Elizabeth R. Reisner
Richard N. Apling
Joanne Bogart
Policy Studies Associates

Second Interim Report from the National Assessment of Chapter 1
Office of Educational Research and Improvement
U.S. Department of Education
July 1986

OR 87-501

Preface

In December 1983, Congress mandated a national assessment of Chapter 1. The requirement, included in the Technical Amendments to the Education Consolidation and Improvement Act (ECIA) of 1981, directed the National Institute of Education (NIE)[1] to conduct independent studies and analyses of compensatory education programs funded under Chapter 1 of ECIA, and to report its findings to Congress by January 1987. The studies and analyses were to address the following topics with respect to Chapter 1:

- services delivered;
- recipients of services;
- background and training of teachers and staff;
- allocation of funds (to school sites);
- coordination with other programs;
- effectiveness of programs on students' basic and higher order academic skills, school attendance, and future education; and
- a national profile of the way in which local educational agencies implement activities described under Section 556(b) of Chapter 1.

The mandate also required consultation with relevant members of the House and Senate education committees. The requirement for the National Assessment of Chapter 1 is reproduced in Appendix A, and a report of its administrative status is contained in Appendix B.

This is the second evaluation of the Federal compensatory education program that Congress has requested. The Education Amendments of 1974 contained the mandate for the previous study. Findings from the resulting Compensatory Education Study, which

---

[1]On October 7, 1985, the NIE became part of the Office of Educational Research and Improvement (OERI) within the U.S. Department of Education (ED).

was also conducted by the NIE, contributed to the formulation of the 1978 reauthorization of Chapter 1's predecessor, Title I of the Elementary and Secondary Education Act (ESEA).

Two significant legislative changes have occurred since 1978. First, in 1981 Chapter 1 of ECIA replaced Title I of ESEA. Chapter 1 retains the purposes of Title I but changed certain administrative features of the program. Second, in 1983 technical amendments to ECIA were enacted in an effort to clarify ambiguities in Chapter 1 and to restore some Title I provisions that had been dropped or changed in Chapter 1. Among these technical amendments, as previously noted, was the requirement for this National Assessment of Chapter 1.

Congress asked that the National Assessment of Chapter 1 provide two interim reports in addition to a final report. The first of the interim reports (Kennedy, Jung, and Orland, 1986) describes the population of students that Chapter 1 is intended to serve--educationally deprived students residing in areas with high concentrations of children from low-income families. This second report reviews and synthesizes evidence regarding the effectiveness of Title I and Chapter 1 programs. Both interim reports draw mainly from data collected in earlier studies or data collection activities. The two interim reports are intended to provide policy makers with a broad perspective from which to view current Chapter 1 programs, which in turn will be described more fully in the third and final report of the National Assessment of Chapter 1.

The third report will draw from the series of studies commissioned specifically for the National Assessment of Chapter 1. These studies are described in Appendix C. The third report will describe:

- The characteristics of Chapter 1 participants;
- The quantity and characteristics of services being provided by Chapter 1;
- Program administration at each level of educational governance; and

- How and why districts make decisions about the selection of schools and students, the allocation of funds among schools, and the design of their Chapter 1 programs.

A particular emphasis of the report will be to identify program practices that have changed or remained the same in the shift from Title I to Chapter 1.

The Chapter 1 Study Team began to implement the National Assessment in the Fall of 1984, after its Study Plan had been reviewed by Congressional staff members in both the Senate and House education committees. Responsibilities for the several components of the National Assessment are distributed among members of the Study Team. Mary Kennedy, Richard Jung, and Martin Orland had primary responsibility for the first interim report. Mary Kennedy and Randy Demaline took the lead in the second interim report. Beatrice Birman, who took over the duties of Director in May 1986, oversaw the completion of the second interim report and is directing work on the final report. Sections within the final report are distributed as follows: Richard Jung is responsible for describing the characteristics of program recipients and patterns of their participation, Gilbert Garcia for describing services, Martin Orland for analyzing administrative practices, and Ron Anson for describing district-level decisions about the program. Paige Russ and Saunders Freeland had primary responsibility for typing this report.

Beatrice F. Birman, Director
National Assessment of the Chapter 1 Program

Ron Anson, Deputy Director
National Assessment of the Chapter 1 Program

Executive Summary

This report reviews findings from several large national data collection efforts designed to estimate the effect of Title I and Chapter 1 on student achievement. The studies and data sources that have been used extensively are described in Appendix D. While each study cited in this report presents its own problems and strengths with respect to its choice of outcome measures and comparison groups, the studies together suggest several broad conclusions about the effectiveness of Chapter 1. These conclusions are summarized below and discussed in the remaining chapters of the report.

Chapter 1 describes Chapter 1 programs and the methods typically used to assess their effects.

Chapter 2 reviews evidence regarding the population of disadvantaged children in general and how their achievement differs from that of the population of all children. The primary finding presented in this chapter is:

1. The achievement of disadvantaged students has improved since 1965, especially in reading, relative to the achievement of the general population.

Chapter 3 describes the one-year effects of Chapter 1-funded programs on students' achievement test scores and describes the differences found across grade levels and between reading and mathematics. Its main findings are:

2. Students receiving Chapter 1 services experience larger increases in their standardized achievement test scores than comparable students who do not. However, their gains do not move them substantially toward the achievement levels of more advantaged students.

3. Students participating in Chapter 1 mathematics programs gain more than those participating in Chapter 1 reading programs.

4. Students in early elementary Chapter 1 programs gain more than students participating in later-grade programs.

5. Evidence regarding program effects on student attitudes toward school is inconclusive.

6. Researchers have not yet developed adequate methods for determining the relationship between program costs and program effects on standardized achievement scores.

Chapter 4 looks beyond a single school year to assess longer-term program effects. Its main findings are:

7. The achievement gap between disadvantaged and advantaged students appears to widen during the summer months.

8. Title I-supported summer programs have not narrowed the gap; however, these programs often were not designed to be academically rigorous.[1]

9. Students who discontinue Title I appear gradually to lose the gains they made when receiving services.

10. Chapter 1 students with very low achievement scores appear to maintain their relative academic positions but not to move ahead. However, the evidence suggests they would have lost ground relative to their peers if they had not received services.

11. No nationally-representative studies have examined the long-term effect of Chapter 1 programs on graduation rates, future education, or adult literacy.

Chapter 5 reviews evidence regarding the advantages and disadvantages of particular project characteristics. Its main findings are:

12. Large-scale studies designed to identify particular project characteristics that improve student achievement test scores have yielded inconsistent or inconclusive results.

13. Researchers have recently identified a number of instructional practices that are likely to increase the achievement of disadvantaged students and that can be used in Chapter 1 programs.

These 13 statements must be interpreted as indicating general trends, which will not necessarily apply to particular projects, schools, or children. Chapter 1 projects vary considerably across states, districts, and even schools within districts. Consequently, no single statement will apply equally well to all activities supported by Chapter 1. Our findings are also somewhat qualified by the age of the data on which

---

[1]These summary statements refer to Chapter 1 except when they are based solely on information collected prior to the enactment of ECIA.

they are based. The most recent data we use to describe achievement gains of Chapter 1 students were reported by state educational agencies for the 1983-84 school year. In addition to these data, we draw heavily on data collected between 1976 and 1979. The use of these earlier data to review the effectiveness of Chapter 1 assumes the program has not changed enough to alter substantially the conclusions that would be reached if a new study were to be undertaken today. Indeed, the authorizing legislation for compensatory education has been generally consistent in its structure and purposes for over 20 years and Chapter 1 programs have become stable entities in many districts. Our final report will provide Congress with details about the current operation of Chapter 1 programs and the extent to which they appear to have changed over time.

Contents

Preface		iii
Executive Summary		vii
1. Introduction		1
	Chapter 1 Services	1
	Measuring the Effects of Chapter 1 Services	3
	Attributing Achievement Growth to Chapter 1 Projects	4
	Program Outcomes Not Addressed in this Report	6
2.	Trends in the Achievement of Disadvantaged Children	7
	Achievement Trends During the 1970s	8
	Recent Evidence of the Achievement of Disadvantaged Students	11
	Summary and Discussion	15
3.	Compensatory Education and Achievement Test Scores	17
	Program Effects on Student Achievement	18
	Program Effects on Student Attitudes Toward School	39
	The Relationship Between Costs and Effects of Chapter 1 Services	40
	Summary and Discussion	41
4.	Sustaining Achievement Through the Summer and Future Years	47
	Extent of Summer Dropoff in Achievement	48
	Impact of Summer School Programs	54
	Retention of Chapter 1 Benefits Obtained During the School Year	55
	Sustaining Achievement Gains Across School Years	60
	Long-Term Program Effects in Areas Other Than Achievement Test Scores	69
	Summary and Discussion	70
5.	Project Characteristics and Student Achievement	75
	Efforts to Identify Effective Projects	76
	Efforts to Identify Effective Features of Chapter 1 Projects	77
	Efforts to Identify Generally Effective Practices Not Limited to Chapter 1	83
	Summary and Implications for Chapter 1	90
Appendices		
A.	Statute Requiring a National Assessment of Chapter 1	A-1
B.	Report of Administrative Status of the National Assessment of Chapter 1	B-1
C.	Studies Commissioned by the National Assessment of Chapter 1	C-1

D. Sources of Data Used in this Report D-1

E. Measures Used in this Report E-1

F. One-Year Achievement Gains Expressed in Normal Curve Equivalents Rather than Percentile Ranks F-1

G. Sample Sizes for Figures and Tables G-1

1. Introduction

In 1965, Congress authorized Federal support for compensatory education through Title I of the Elementary and Secondary Education Act (ESEA). Since then, there have been numerous studies and reviews of the effectiveness of programs supported by that legislation. The conclusions from these studies and reviews have varied, depending on when, where, and how the studies were conducted. This report examines and summarizes existing data on the effects on students' academic achievement of ESEA Title I and its successor, ECIA Chapter 1. Appendix D describes the studies and data bases most heavily relied upon for this report.

Chapter 1 Services

The effects of Chapter 1 have been difficult to assess at the national level, in part because Chapter 1 does not require a particular instructional program. Instead, it permits districts to design programs they believe will promote the educational development of their particular population of students. Consequently, districts and even schools within districts differ in the grade levels they serve, the procedures they use to select students, the services they provide, and the administrative strategies they use to orchestrate those services.

Figure 1.1 indicates that 75 percent of the participating students receive instructional services in reading, and that nearly half of all participants receive instruction in mathematics. Chapter 1 students also receive instruction in language arts and in special subjects such as English as a second language, and they receive support services such as attendance, guidance, health, and nutrition assistance. Figure 1.1 does not indicate, however, the variety of services that may be provided within a single service category such as reading or mathematics. A student may receive Chapter 1 reading services, for instance, for 20 minutes three days a week or

Figure 1.1

Percent of Chapter I Students Receiving Instructional and Non-Instructional Services.* 1983-84

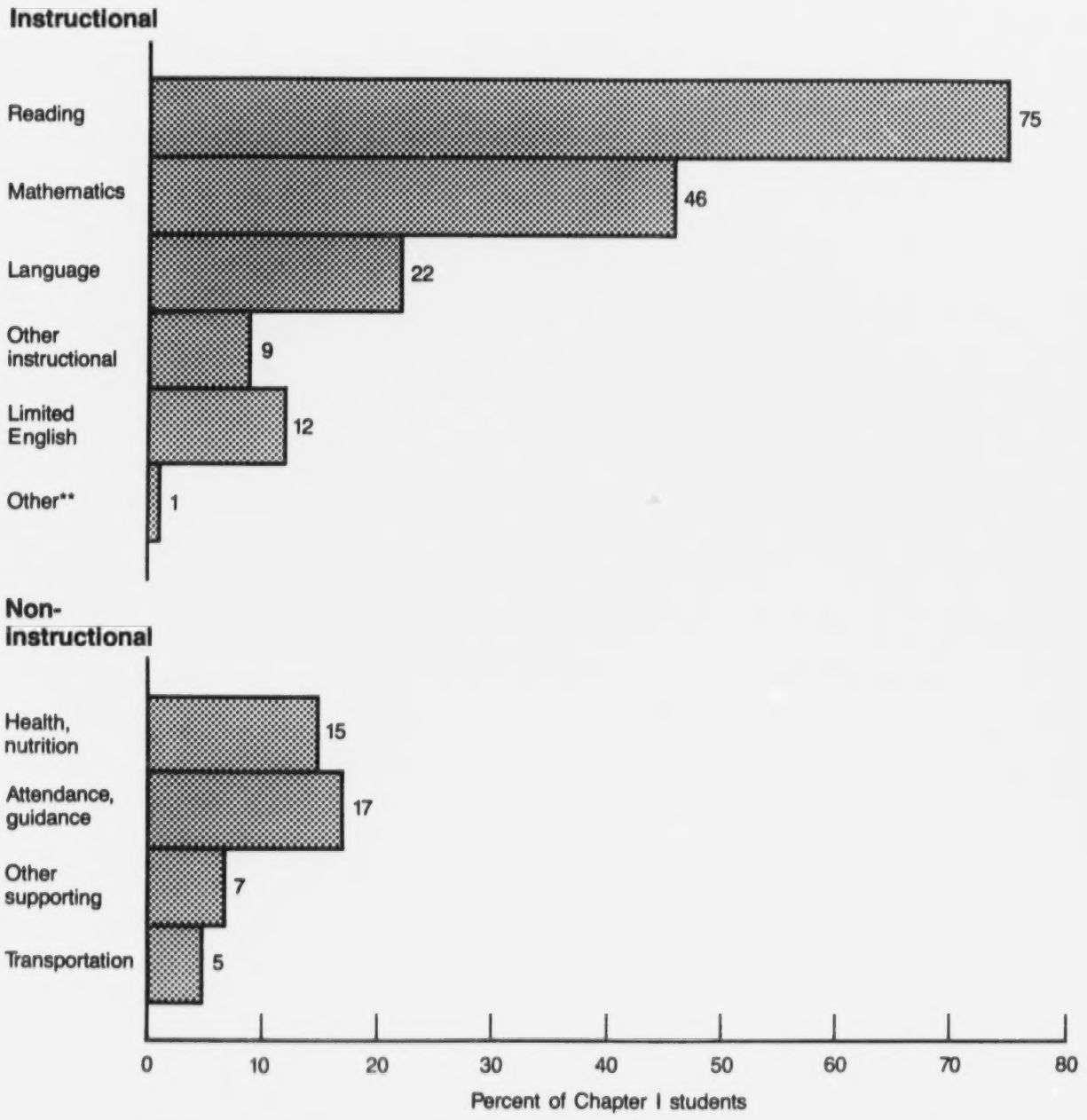


Figure reads: Seventy five percent of all students enrolled in Chapter I programs received supplementary reading instruction during the 1983-84 school year; 46 percent received supplementary mathematics instruction.

*Total number of students = 4.85 million. If children receive more than one service, they are counted in each subject in which they are enrolled.

**Includes vocational instruction and special services for handicapped students.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology, 1985.

for two hours every day.[1] These services may be provided in the child's regular classroom or in another setting; they may be provided by teachers, teacher aides, reading or math specialists or parents; and they may rely on the same instructional materials as the regular classroom or use different materials. Even when services are all labeled "reading," the particular reading skills emphasized can vary dramatically.

Measuring the Effects of Chapter 1 Services

Studies of Chapter 1 have usually responded to this diversity by limiting their scope to reading and mathematics programs. Such a focus results in more manageable evaluations, and covers the most commonly provided services. But it also precludes evaluation of the impact of other Chapter 1 services.

Studies of Chapter 1 have also limited their own scope by measuring the impact of Chapter 1 programs mainly with standardized achievement tests. These tests cover the range of academic content taught to the general student population. They are not intended to measure the specific content taught by any one instructional program. Consequently, they will measure achievement in areas that were never intended to be taught, and will fail to document some of the skills that students learn as a result of their participation in a program. This mismatch between the content of the test and the content of the curriculum is unavoidable, because standardized tests are not designed to measure changes attributable to specific programs but instead to provide school districts and others with a general sense of what students know, relative to the population in general. Yet because of this content mismatch, when standardized achievement tests are used to measure the growth of a particular student and to

---

[1]Services provided for more than 25 percent of the school day require a contribution of resources from the school district.

attribute that growth to a particular program, the tests will generally underestimate the amount of program-related growth students have actually experienced during the year.

Some educators have argued that student outcomes such as attitude toward school, motivation, and high school graduation are more meaningful indicators of a program's success than are changes in achievement test scores. Indeed, the long-term consequences of changes in achievement test scores is not always clear. Thus, while they are valuable indicators of student learning, achievement test scores do not reflect the long-term consequences of Chapter 1 program participation, and they only imperfectly measure the basic reading and mathematics knowledge that they are designed to assess.

Attributing Achievement Growth to Chapter 1 Projects

Several strategies have been used to estimate the impact of Chapter 1 services. One is to compare the achievement of disadvantaged students to the achievement of advantaged students, and to determine whether the gap between these two groups is reduced over time. Some analysts have used data from the National Assessment of Educational Progress in this way, and their findings are reviewed in this report. This strategy does not take into account which students actually receive program services and which do not. Consequently, it cannot determine the extent to which Chapter 1-funded programs contributed to any observed population changes.

It would be possible to alter these comparisons to include only disadvantaged students who actually have participated in the program, and determine whether the gap between this group and the population of advantaged students changes as participants receive services. This approach would describe the extent to which the achievement levels of Chapter 1 participants came to resemble that of their more advantaged peers. But because advantaged and disadvantaged students may have different learning rates

anyway, this approach would not indicate how much larger or smaller the observed changes were than would have been expected if the disadvantaged students had not received Chapter 1 services.

The strategy preferred by evaluators is to compare the achievement of students receiving Chapter 1 services to comparable disadvantaged students not receiving Chapter 1 services. But this approach is difficult to implement. It requires the evaluator to identify and test a group of students who do not receive Chapter 1 services, but who are comparable to those who do. Yet if Chapter 1 funds are distributed appropriately, there should be very few schools or students comparable to those receiving Chapter 1 services within a given school district. Chapter 1 requires school systems to provide their Chapter 1 services in schools with the greatest need and to students within those schools who are themselves among those most in need. If evaluators select comparison groups from within Chapter 1 schools, they generally select students who achieve at higher levels than the Chapter 1 participants. On the other hand, if evaluators select comparison students from schools that do not offer Chapter 1 programs, they may find students with comparable achievement scores but with superior educational environments.[2] The more the comparison group differs from the group of students receiving Chapter 1 services, the more difficult it is to use its achievement as a benchmark for estimating the effects of Chapter 1. Despite the difficulty of implementing this approach, it has the advantage of estimating the achievement growth that Chapter 1 participants would have demonstrated if they had not been served.

---

[2]In our first report to Congress, we showed that student achievement was adversely affected by attending schools with high concentrations of poverty. Generally speaking, non-Chapter 1 schools have unusually low concentrations of poverty, and consequently are likely to provide more beneficial learning environments for students.

Program Outcomes Not Addressed in this Report

The analyses examined in this report tell us a great deal about the impact Chapter 1 has had on student achievement. But they do not address all of the student outcomes of interest to Congress. In its mandate for this study, Congress asked for information about " . . . effectiveness of programs on students' basic and higher order academic skills, school attendance, and future education . . ." Because this report used only existing data and because these data were mainly standardized achievement test scores gathered within a year or two of students' receipt of Chapter 1 services, the report does not address three of Congress' concerns. First, standardized tests do not separately measure higher-order skills such as written composition and abstract problem solving and consequently do not permit us to distinguish the potential contributions Chapter 1 may have made in these areas. Second, no national data base contains evidence regarding the school attendance of Chapter 1 students. And finally, no national study has followed students for more than two years after they leave Chapter 1, so that it is not possible to ascertain the program's impact on students' future education.

There is some evidence that students' outcomes in such areas as school attendance and future education differ from their test score improvements. In studies of early childhood programs (Smith, 1985), student test scores rose while students received services and later declined, yet other indicators of educational success continued to show evidence of program benefit. Such a pattern may not appear for Chapter 1 students, since Chapter 1 services are generally more limited in scope than are preschool services. Nevertheless, these patterns serve as a reminder that, while the achievement data presented here are important, they do not present a complete picture of the potential effects associated with Chapter 1 programs.

2. Trends in the Achievement of Disadvantaged Children

Central Finding

1. The achievement of disadvantaged students has improved since 1965, especially in reading, relative to the achievement of the general population.

This chapter reviews evidence regarding achievement trends in the population of school-age children. It contrasts changes that have occurred over time among different subgroups of the population, with particular emphasis on those groups most likely to have received Chapter 1 services.

Chapter 1 is designed to support services specifically for low-achieving students who attend schools with high concentrations of children from poor families. The legislation provides extensive guidance regarding the selection of both schools and students to participate in local programs. Approximately 5 million students participate in Chapter 1 each year, about 11 percent of the entire student body (Carpenter and Hopper, 1985). Of these, approximately 40 percent are new to the program each year, while a comparable group who received services the preceding year are removed from the program (Carter, 1984). Because a slightly different group receives services each year, the proportion receiving services all together is larger than the proportion receiving services at any one time. By the time students in some districts reach fourth grade, for instance, an estimated 25 percent have received services at some time during their school careers.[3] In districts and schools that serve larger proportions of students within each school year, the total percent served over time would be even larger. Given the legislated rules regarding participation, it is reasonable to suppose that these students represent the most educationally-deprived students within each

---

[3]This estimate is based on multi-year participation data in two districts--St. Louis, Missouri and Lincoln, Nebraska. For details, see Pfannensteil, 1986.

district.[4]

The amount of services these students receive is also highly variable. A small fraction of these fourth grade children have received Chapter 1 services daily for the entire five-year period; others have received services throughout one or two school years, and still others received a few months or even just a few weeks of special assistance. Furthermore, all of these students received, throughout the entire five-year period, their regular school program. If we were to examine the entire population of fourth graders, the particular contribution Chapter 1 has made to their educational achievement as a group would be extremely difficult to define, in part because the nature and extent of their participation in Chapter 1 would have been so variable, and in part because it would not be possible to separate the benefits of these services from the benefits of their regular full-time educational program.

Without knowing which students received services, or when or for how many years they received services, the most general way to examine Chapter 1's effects on student achievement is to examine changes in the achievement of all disadvantaged students, and to overlook considerations of the nature or extent of their actual participation in the program. Such an examination cannot indicate the specific effects of Chapter 1, of course, but it can indicate important trends in the achievement of the population that Chapter 1 is intended to serve.

Achievement Trends During the 1970s

The National Assessment of Educational Progress (NAEP) gathers nationally-representative achievement data on 9, 13, and 17 year-old students. Although NAEP does not identify all categories of students who might be designated as educationally deprived, it does provide several groupings relevant to our analytic purposes. These

---

[4]In our first report to Congress, however, we indicated that many participating schools do not serve particularly large proportions of low-income children.

include (1) low-achieving students, defined as those in NAEP's lowest achievement quartile, (2) students attending schools located in disadvantaged urban communities; and (3) minority students. The first two of these groups are the specific focus of Chapter 1's student and school selection procedures and its funding formula. The third is not specifically targeted; however participation data indicate that a relatively large proportion of minorities do participate in local Chapter 1 programs.

In 1984, the director of NAEP delivered testimony to Congress regarding changes in the academic progress of many of these students (LaPointe, 1984). His findings are presented in Table 2.1. The table entries indicate changes over time in the average percent of test items that students answered correctly. Asterisks mark the entries that are statistically significant.

Table 2.1 indicates several important patterns. First, improvements were larger among students in the lowest quartile (the lowest-achieving 25 percent of the total student population) than among those in the highest quartile, especially in reading. Second, black students tended to demonstrate more and larger changes than white students. Third, 9-year-olds gained more than 13- and 17-year-olds, especially in reading and mathematics; and finally, students improved more in reading and mathematics than in science.[5]

Not all of these patterns have been corroborated in other data bases. However, using these data, LaPointe (1984) and Riddle (1984) have argued that the patterns of increases in NAEP achievement data parallel patterns of Federal financial support to schools. For instance, most Federal funds support programs for younger students, and

---

[5]Table 2.1 also shows some important patterns of <u>decreases</u> over time. For instance, the top quartile of white students at all age levels decreased their percent correct in mathematics and science items during this period of time. This is clearly a pattern worth attending to, though its relationship to Chapter 1 services is unclear.

Table 2.1

Change in Percent of Correct Items by Subject, Race, Age, and Achievement Quartile During the 1970s,** NAEP

		Black Students			White Students		
Quartile	Age	Reading	Science	Mathematics	Reading	Science	Mathematics
Lowest 25% of Students	9	8.4*	−0.7	2.9*	4.6*	1.7*	−0.5
	13	3.5*	1.3	2.6*	1.5*	2.0*	0.3
	17	1.1	−0.5	1.6*	−1.7*	0.7	−1.8*
Highest 25% of Students	9	3.0*	1.1	2.6*	1.2	−2.4*	−3.3*
	13	2.5*	−0.5	−2.5*	0.4	−4.1*	−3.2*
	17	−1.1	−3.9*	−5.5	−0.3	−4.2*	−4.3*

Table reads: Among students in the lowest achieving quartile, the percent of reading items answered correctly by 9-year-old black students increased 8.4 percentage points. The percent of reading items answered correctly by 9-year-old white students in the same group increased 4.6 percentage points.

*Denotes figures that are statistically significant at the 5 percent confidence level — i.e., given the range of scores and the size of the relevant populations, there is a probability of less than 5 percent that the indicated change in average scores has occurred simply by chance.

**The specific time intervals vary with the subject. In reading, the test points were 1970, 1974, and 1979. In mathematics, they were 1972, 1977, and 1981. In science, they were 1969, 1972, and 1976.

Source: W. Riddle, *Achievement Score Trends and Federal Involvement in Elementary and Secondary Education: An Exploration of Their Relationships.* Report 84-627EPW. Washington, D.C.: Congressional Research Service, The Library of Congress. May 3, 1984.

this was the age group showing the largest gains. Second, Federal funds focus on low-achieving students, which is the population for whom the greatest improvements were observed. Third, Federal funds provide services to higher proportions of minority than nonminority students, and minority students have shown greater achievement improvements than nonminority students. Finally, Federally-supported services focus on core subjects such as reading and mathematics, and the achievement increases have occurred in these areas rather than in science. According to both authors, the general pattern of test performance changes over time seems to parallel the pattern of Federal education investments. However, both authors discussed Federal programs in general, rather than Chapter 1 in particular, and Federal investments in other programs were at their peak during the period covered by these analyses. Furthermore, any number of other social changes that occurred during this period of time could have influenced these patterns. Consequently, it is difficult to estimate the extent to which the patterns of improvement indicated by NAEP derive from the particular children who received Chapter 1 services.

Recent Evidence of the Achievement of Disadvantaged Students

Since LaPointe and Riddle presented their data in 1984, two new sources of evidence on achievement trends have become available: (1) a new NAEP report on reading achievement summarized data from 1971 to 1984 (NAEP, 1985), and (2) a report by the Congressional Budget Office (CBO) summarized achievement trends found in a number of testing programs (Koretz, 1986).

Figure 2.1 presents the recent NAEP findings regarding reading achievement among black, Hispanic, and white students. All three groups show evidence of improvement during this time interval, but black and Hispanic students started at much lower levels of achievement and showed much greater change over time. For instance,

Figure 2.1

Trends in Average NAEP Reading Proficiency Scores* for Black, Hispanic, and White Students

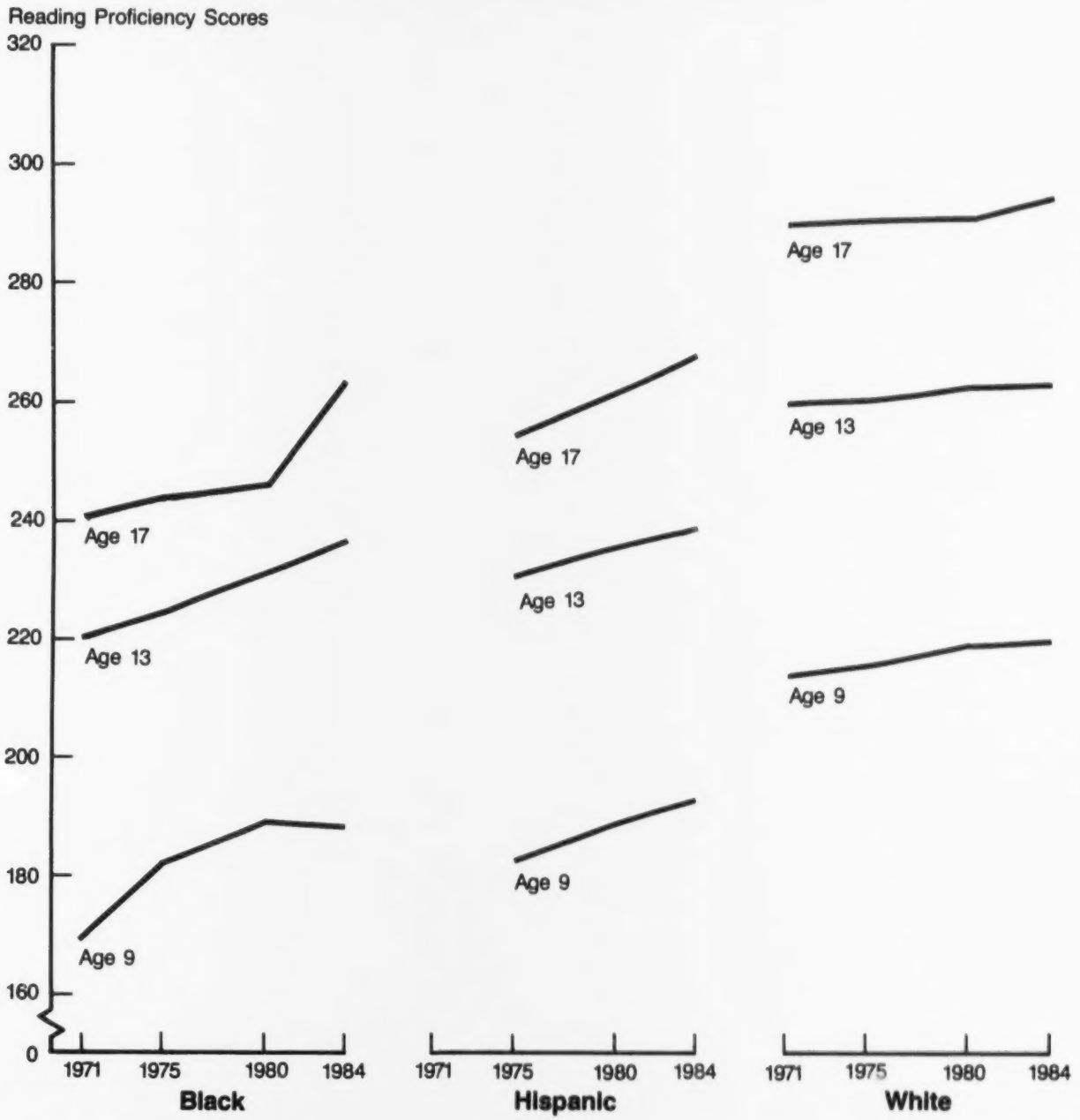


Figure reads: The average reading score of black 9 year olds was approximately 169 in 1971 and 188 in 1984, a gain of 19 points. The average score of white 9 year olds was 214 in 1971 and 220 in 1984, a gain of 6 points.

*These scores are derived from item response theory. Based on a scale that ranges from 0-500, these scores provide a common scale on which comparisons can be made for different age and test groups. Scores on the scale relate to five levels of proficiency: rudimentary (150), basic (200), intermediate (250), adept (300), and advanced (350).

Source: National Assessment of Educational Progress, *The Reading Report Card, Progress Toward Excellence in our Schools: Trends in Reading Over Four National Assessments, 1971-1984.* Princeton: Educational Testing Service, 1985.

black 9-year-olds scored around 20 points higher in 1980 and 1984 than they did in 1971. The reading performance of black students in 1984 still lags far behind white students, but the size of the gap has been reduced considerably in comparison to the gap in 1971. The patterns of minority achievement found in 1984, then, are still consistent with those originally observed by LaPointe.

These NAEP analyses do not examine changes in each achievement quartile and so cannot be compared to the earlier findings on that dimension. However, they do indicate achievement trends among students living in various types of communities. The NAEP findings on this dimension are shown in Figure 2.2. This figure indicates that achievement has improved among students residing in rural and disadvantaged urban areas more than it has in advantaged urban areas,[6] with the most substantial improvements among 9-year-olds residing in disadvantaged urban areas.

Interest in these and other achievement patterns prompted the CBO to examine achievement trends evident in NAEP and in other data bases (Koretz, 1986). The CBO analysis was primarily concerned with the overall national decline and subsequent upturn in student achievement test scores. It enables us to place the NAEP achievement patterns in a larger context. CBO found, for instance, that while achievement scores on a wide variety of tests had been declining through the decades of the 1960s and 1970s, the trend actually began to reverse itself with children born in approximately 1963. These children entered school in about 1968, three years after the passage of the Elementary and Secondary Education Act of 1965, and were nine years old in 1972. CBO also found that the decline was most severe among older

---

[6]NAEP describes its residential areas according to employment statistics. For instance, a disadvantaged urban area is one with a population of at least 200,000 and with an unusually small proportion of managerial and professional workers. Rural areas are defined only by population density, and could include advantaged as well as disadvantaged families.

Figure 2.2

Trends in Average NAEP Reading Proficiency Scores* by Type of Community

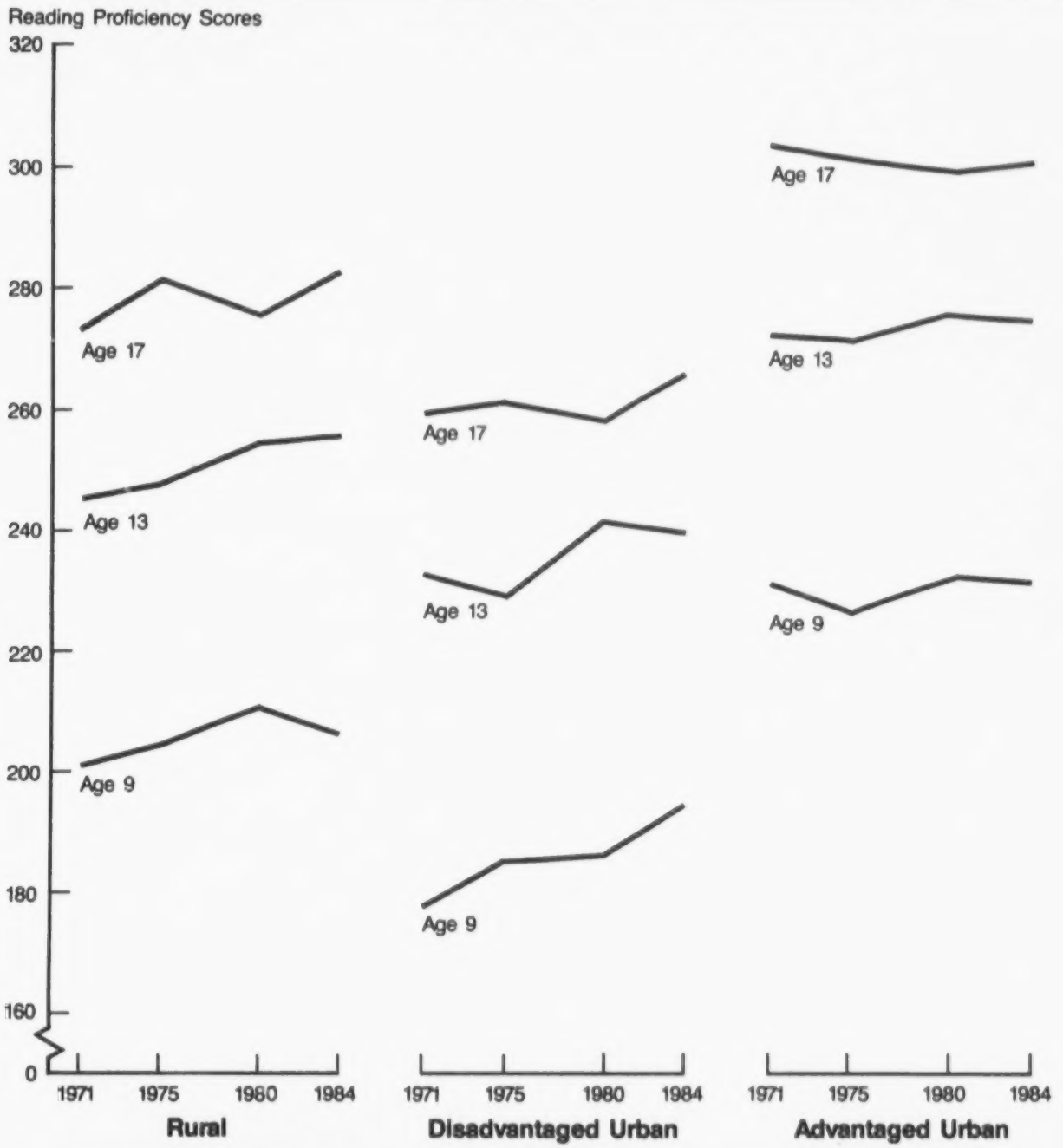


Figure reads: The average reading proficiency score for 9 year olds in rural communities was approximately 201 in 1971 and 206 in 1984, a 5 point increase. In disadvantaged urban communities, 9 year olds scored 178 in 1971 and 194 in 1984, a 16 point gain. In advantaged urban settings, 9 year olds scored 231 in both 1971 and 1984.

*These scores are derived from item response theory. Based on a scale that ranges from 0-500, these scores provide a common scale on which comparisons can be made for different age and test groups. Scores on the scale equate with five proficiency levels: rudimentary (150), basic (200), intermediate (250), adept (300), and advanced (350).

Source: National Assessment of Educational Progress, *The Reading Report Card, Progress Toward Excellence in our Schools: Trends in Reading Over Four National Assessments, 1971-1984.* Princeton: Educational Testing Service, 1985.

students and least severe among younger students. In fact it found that scores of students in the upper elementary grades are at their highest level in three decades. Third, CBO found the decline was most severe on measures that required students to use higher-order thinking skills such as problem solving or drawing inferences, and least severe on measures that require more basic skills such as recalling facts or computing. Finally, CBO found that the general decline in achievement test scores was less apparent among minority students than among other groups, while the upturn was particularly strong among minority students, especially those in elementary and early secondary grades. The CBO analysis did not find, however, clear evidence of differences in improvements between low-achieving and high-achieving students, and it did not find clear differences in the long-term trends of reading and mathematics versus science.

Summary and Discussion

The central finding reported in this chapter is:

1. The achievement of disadvantaged students has improved since 1965, especially in reading, relative to the achievement of the general population.

Although analyses of population trends cannot be used to infer direct programmatic effects, such analyses provide important information about the achievement of students who are intended to benefit from Chapter 1 services. As we have seen, it is possible that a substantial portion of educationally-deprived children have received some amount of Chapter 1 services at some point in their educational careers. Their participation in Chapter 1 could have contributed to the observed improvements in the educational progress of disadvantaged and minority students.

On the other hand, many events in the past decade and a half could account for the changes reported by NAEP and CBO. School districts have tried to desegregate their schools; State and local educational agencies have increased their focus on basic

skills and their overall spending on educational services for disadvantaged students; minority groups have become more self conscious and politically active; and teachers have become more aware of their obligations to low-achieving and minority students.

The most substantial improvements during this period occurred among black and Hispanic students, regardless of residence, and among students who reside in disadvantaged urban areas, regardless of their race or ethnicity. Thus, trends in student achievement suggest that the nation is improving its education of disadvantaged students. To identify and measure the specific impact of Chapter 1 programs on achievement, we turn now to a review of available national data contrasting students who actually participated in Chapter 1 programs with comparable students who did not.

3. Compensatory Education and Achievement Test Scores

Central Findings

Program Effects on Student Achievement

2. Students receiving Chapter 1 services experience larger increases in their standardized achievement test scores than comparable students who do not. However, their gains do not move them substantially toward the achievement levels of more advantaged students.

3. Students participating in Chapter 1 mathematics programs gain more than those participating in Chapter 1 reading programs.

4. Students in early elementary Chapter 1 programs gain more than students participating in later-grade programs.

Program Effects On Other Aspects of Student Development

5. Evidence regarding program effects on students' attitude toward school is inconclusive.

Costs Relative to Effects

6. Researchers have not yet developed adequate methods for determining the relationship between program costs and program effects on standardized achievement scores.

This chapter focuses not on the population of disadvantaged children in general but on students who actually receive Chapter 1 services. It presents evidence describing the average effects of Chapter 1 on participating students and how these effects vary across grade levels and subject areas. The chapter also examines briefly Chapter 1 effects on nonacademic areas of student development and the relationship between program costs and program effects.

Program Effects on Student Achievement

To address the question of whether Chapter 1 programs increase students' educational achievement, we rely primarily on data from two sources: the Title I/Chapter 1 Evaluation and Reporting System (TIERS) and the Sustaining Effects Study.

TIERS Achievement Data

Developed in the mid-1970s, TIERS provided a framework within which State and local educational agencies would be required to report comparable data to the U.S. Department of Education regarding their Chapter 1 projects and the students they serve. The system included standardized procedures for school districts to use in reporting the number of students receiving various Chapter 1 services and in measuring the impact of these services on student achievement. TIERS permitted the aggregation of project data to State and then to national levels, thus enabling the development of a national summary of achievement test scores for Chapter 1 participants. The ECIA, enacted in 1981, eliminated the requirement that State and local educational agencies implement these or any other standard evaluation procedures, though many State and local agencies continue voluntarily to use the TIERS procedures.[7]

Though Chapter 1 services are offered to students ranging from pre-kindergarten through twelfth grade, the services are concentrated in the elementary grades, with nearly 70 percent of Chapter 1 participants in grades one through six. Figure 3.1 displays the percentage of Chapter 1 students who receive services in each grade level, pre-K through 12. It indicates that Chapter 1 students are roughly equally distributed among grades one through six, and that progressively fewer students are served in each successive grade.

---

[7]Local projects are still required to evaluate their programs' impact on students annually, but they need not use standardized procedures to do so.

Figure 3.1

Percent of Chapter I Students Receiving Services by Grade, 1983-84

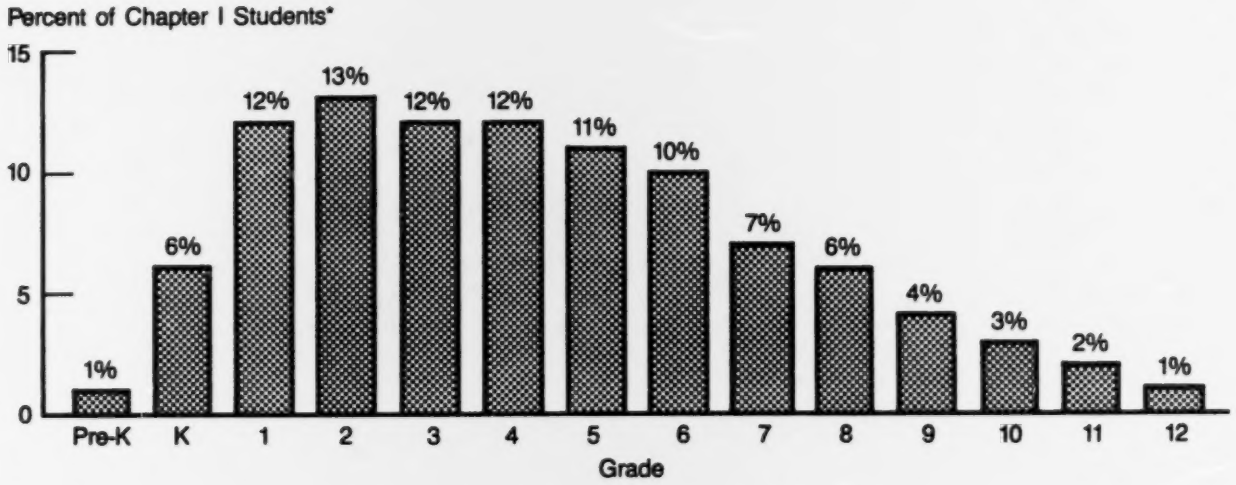


Figure reads: One percent of all Chapter I students served in the 1983-84 school year were enrolled in pre-kindergarten and six percent were enrolled in kindergarten.

*Total number of students = 4.85 million

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology, 1985.

Figure 3.2 summarizes the average achievement levels of students tested under TIERS in 1983-84 for reading and mathematics.[8] The slanting lines in Figure 3.2 represent the change in average scores, expressed as percentile ranks, for students receiving Chapter 1 services during the 1983-84 school year in grades two through twelve.[9] A percentile rank is a form of test score indicating the percent of all students nationwide who scored below that achievement level. The average student nationwide would achieve at the 50th percentile rank at the beginning of the year and at the end of the year as well. In Figure 3.2, the average second-grade Chapter 1 student performed at the 29th percentile rank in reading in spring 1983[10]; higher than 29 percent of all second-grade students tested in national norming samples. In spring 1984, after having received reading services supported by Chapter 1, this average Chapter 1 student's score was at the 31st percentile, two percentile ranks higher than a year before. Thus, the average second grader participating in Chapter 1 gained enough during the school year to surpass two percent of the nation's second-grade students by the end of the school year.[11]

---

[8]Some States reported for only a few grades or a few districts, and some submitted no report. It is difficult to know the extent to which State decisions to submit or not to submit particular data elements influence the validity of national averages. However, we do know that TIERS data show very similar national patterns from year to year, even though State participation fluctuates and State-specific patterns fluctuate.

[9]TIERS does not provide for the collection of achievement data from first-grade students, because it is not feasible to test these students in the spring preceding their enrollment in first grade. First grade data are also generally recognized as less reliable.

[10]Districts have the option of testing students each fall and each spring, and thereby measuring school-year gains, or of measuring students only once a year--each spring or each fall -- and thereby measuring gains over the entire 12-month year. The two methods yield remarkably different results, which are discussed in Chapter 4. The analysis in this chapter uses only the annual test schedule.

[11]These data include the full range of Chapter 1 students -- successful Chapter 1 students, who "graduate" and did not return to the program the next year, as well as the least successful, who may have returned to the program for several more years, or who may have eventually been transferred to special education.

Figure 3.2

Changes in Percentile Ranks* for Chapter I Students in Reading and Mathematics, 1983-84**

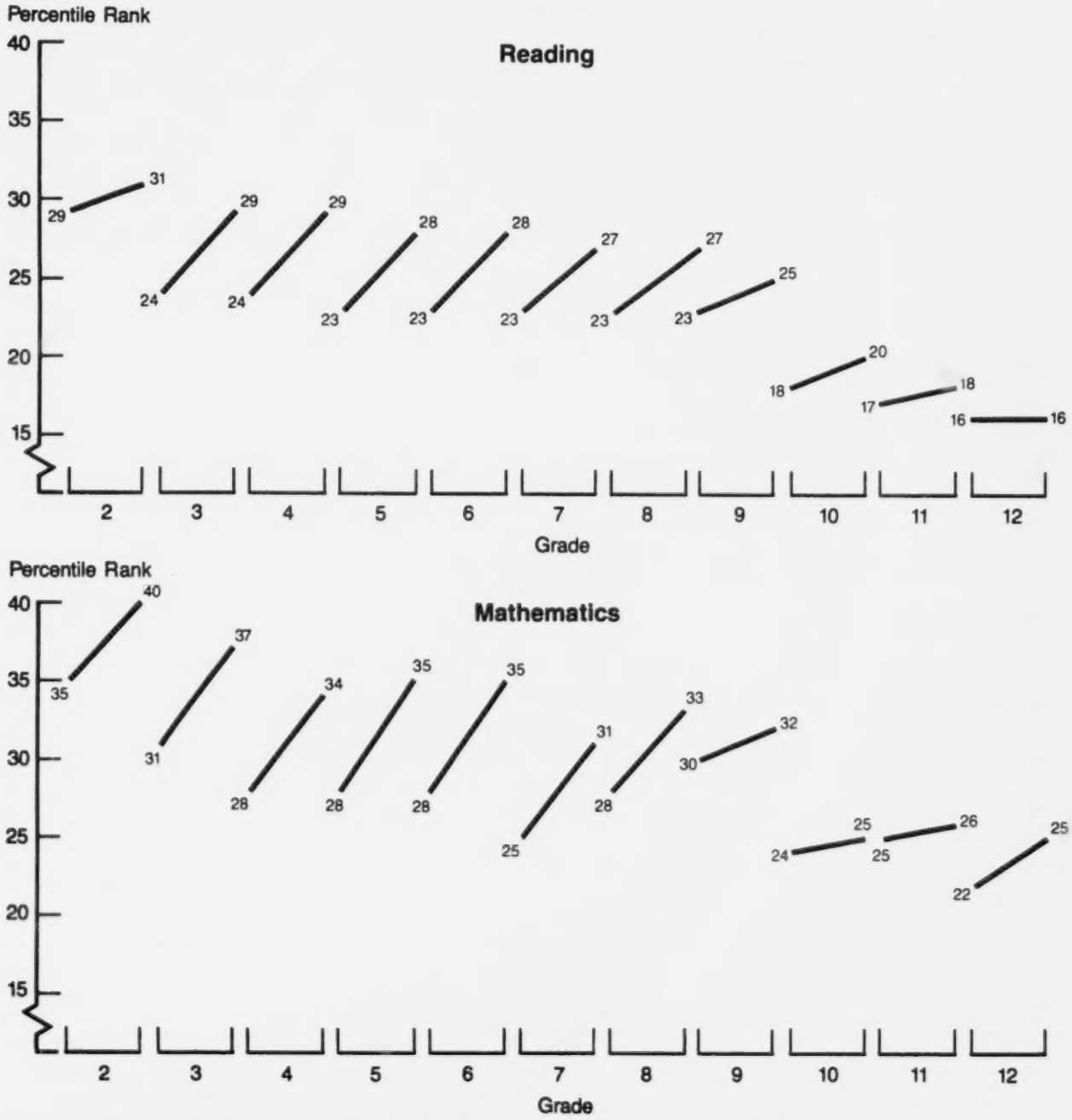


Figure reads: From spring 1983 to spring 1984, the percentile rank of second grade students who received Chapter I reading instruction increased from the 29th percentile to the 31st, while the rank of 12th grade students remained constant at the 16th percentile.

*Changes in percentile ranks were calculated by first determining all averages in normal curve equivalents (NCEs), a standardized scale score metric, and then converting these averages to percentile ranks. See Appendix E for definitions of the measures and Appendix F for the comparable figure presented in NCEs.

**The number of students included in these analyses varied by grade level and subject; see Appendix G for details.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology, 1985.

Figure 3.2 suggests that students entering Chapter 1 reading programs tended to score at lower percentile ranks than did students entering Chapter 1 mathematics programs. The starting percentile ranks for average reading achievement range from 16 to 29, while the starting levels in mathematics range from 22 to 35. It also shows that students tended to enter both reading and mathematics programs with relatively higher scores in elementary grades and relatively lower scores in secondary grades. Students entering second-grade reading programs had average scores at the 29th percentile rank, while those entering later elementary reading programs averaged at the 23rd or 24th percentile rank, and those entering secondary reading programs averaged at the 16th to 18th percentile rank. A similar pattern, though moderated somewhat, appears among students participating in Chapter 1 mathematics programs.

With respect to their achievement at the end of the school year, nearly all (21 out of 22) of the changes shown in Figure 3.2 indicate an upward movement in percentile ranks of students' average scores. The size of these increases is often only a few percentile ranks, and Chapter 1 students' achievement at the end of the school year was still far from the median, or 50th percentile rank. In general, students receiving Chapter 1 services in mathematics demonstrated slightly larger gains than those of students receiving Chapter 1 reading services. Their improvements ranged from one to seven percentile ranks, whereas reading students increased from zero to five percentile ranks. In both subjects, the size of the average annual gain is smaller for older students. Whereas elementary students receiving Chapter 1 services tended to improve by four or five percentile ranks in reading and six or seven percentile ranks in mathematics, secondary students tended to improve by only one or two percentile ranks in either subject. Thus, Chapter 1 students in the later grades started with a greater educational disadvantage at the beginning of the school year and gained less from their participation in Chapter 1 programs than did elementary school children.

Key Assumptions Underlying the TIERS Analyses

The evidence presented here is based on nationally standardized achievement tests. Gains of Chapter 1 students were not compared to gains of comparable students, but instead were assessed relative to the entire distribution of scores. For the evidence to be meaningful, analysts make two important assumptions. First, they assume that students normally progressing through school without Chapter 1 assistance would maintain their same relative rank at the end of the school year as at the beginning. This is called the "equi-percentile" assumption (Linn, Dunbar, Harnisch, and Hasting 1982). Second, analysts assume that students not participating in Chapter 1 who score at the same percentile rank as Chapter 1 students are in fact comparable to Chapter 1 students.

With respect to the first assumption, available evidence suggests that percentile ranks are not particularly stable over time. Students may move up or down, relative to their peers, during a school year. We do know that there is a strong relationship between student test scores on one occasion and student test scores on other occasions, thus suggesting that scores do not change radically over time.[12] If individual ranks were as likely to rise as to fall, the changes would cancel one another and would not interfere with our interpretation of the data. Later, we will present evidence suggesting that lower-achieving students tend to fall farther and farther behind their higher-achieving counterparts.

---

[12]It is also known that students with lower test scores will show larger growth than other students, a phenomenon known as "regression to the mean". Some researchers argue that, because of regression to the mean, these estimates of student growth are artificially inflated. However, the TIERS system requires districts to select Chapter 1 students with a different test than they use as a pretest, a practice that should reduce the inflation of growth estimates caused by regression to the mean. Further, the patterns displayed here indicate that groups that begin with relatively lower scores also tend to gain less than their higher-achieving counterparts.

With respect to the second assumption--that all students scoring at a given percentile rank are comparable--we expect that students selected to receive Chapter 1 services are more disadvantaged than those not selected, regardless of the comparability of their test scores. Teachers are permitted to use their personal judgment in selecting students to receive program services. If they exercise their judgment in a manner consistent with Chapter 1 requirements, they will select from among comparably-scoring students those whom they believe to be most in need of Chapter 1 services; that is, students for whom they believe the test scores overestimate actual achievement. If the real achievement of Chapter 1 students is lower than their starting scores indicate, then their real gains would be larger than those shown here.[13]

These problems in interpretation arise in part because the TIERS procedures do not include tests of students comparable to Chapter 1 participants to see how their achievement actually changes over the year, but instead rely on the hypothetical progress of students who began the school year with the same achievement scores.[14] This deficiency was addressed in the Sustaining Effects Study, which tested all students in its sample of Title I elementary schools as well as some students in elementary schools without Title I, in order to directly measure the performance of non-participating students.

---

[13]This analysis assumes districts place their lowest-achieving students in Title I programs. In some States and districts, the lowest-achieving students are placed in State or local programs, so that relatively higher-achieving students are placed in Chapter 1. Even in these districts, however, students selected for service from among the remaining unserved children should still be the lowest-achieving students available.

[14]The TIERS actually included several evaluation models which districts could use, one of which included a local comparison group. However, it has only rarely been used by districts. The vast majority of districts prefer the model which relies on the norm group for comparison.

The Sustaining Effects Study

The Sustaining Effects Study examined Title I programs in grades one through six in a representative sample of schools. The researchers measured student achievement before and after one school year (1976-77) in 243 schools and then followed a subset of students[15] over two additional school years (1977-79).

Using the same format as that employed to display the TIERS data, Figure 3.3 translates Sustaining Effects Study findings into the percentile ranks of the average Title I students in 1976-77. Unlike the earlier displays of TIERS data, this figure documents students' average entering achievement in the fall of each school year, rather than in the spring of the preceding school year. From Figure 3.3 we see that the average second-grade student entered Title I reading programs in the fall with a percentile rank of 26 and finished in the spring with a rank of 26; comparable percentile ranks for participants in Title I mathematics programs were 28 and 29.

Although the Sustaining Effects Study data were collected eight years earlier than the TIERS data, they suggest that students served under Title I were comparable in their starting achievement levels to those served more recently, but that the program benefits were somewhat less. Whereas 1983-84 Chapter 1 students improved their average reading score by as much as three to five percentile ranks in elementary grades, comparable 1976 student averages improved by less than two percentile ranks per grade level. Aside from this important difference in the average improvements, Figure 3.3 shows patterns of findings that are remarkably similar to those shown in Figure 3.2. Both figures indicate that students entered reading programs at slightly lower percentile ranks than they entered mathematics programs with, and that students served in earlier grades tended to enter at higher percentile ranks than those served

---

[15]The original sample, on which our findings are based, included 120,000 students across six grade levels.

Figure 3.3

Changes in Percentile Ranks for Title I Students in Reading and Mathematics, The Sustaining Effects Study, 1976-77*

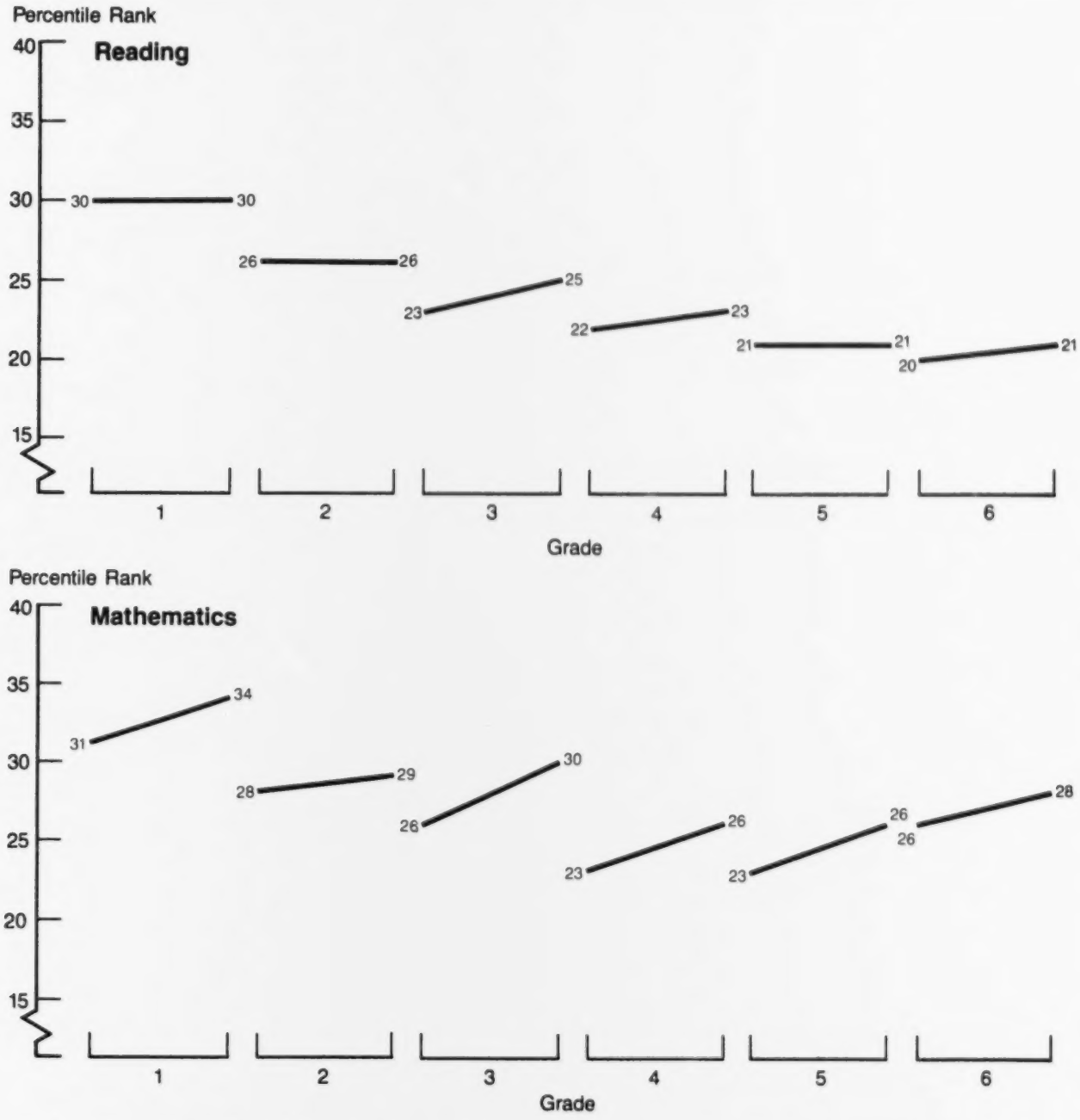


Figure reads: From the fall to the spring testing, 3rd grade students enrolled in Title I reading moved from the 23rd percentile rank to the 25th percentile.

*Percentile ranks presented are based on scores from a fall-spring testing cycle in contrast with the spring-spring cycle used for TIERS data in Figure 3.2. Changes in percentile ranks were calculated by first determining all averages in a standardized scale score metric, and then converting these averages to percentile ranks . See Appendix E for definitions of the measures and Appendix F for the comparable figure presented in NCEs. The number of students included in these analyses varied by grade level and subject; see Appendix G for details.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, *Report 10: Compensatory Services and Educational Development in the School Year.* Santa Monica, CA: System Development Corp., 1981.

in later grades. Further, students participating in mathematics programs tended to enjoy slightly more relative improvement than did students participating in reading programs.

Because the Sustaining Effects Study researchers collected achievement data on all students in their sampled schools, they were able to compare the achievement gains of Title I children to those of comparable students who did not receive services. In Figure 3.4, the solid lines indicate the percentile ranks of students who received Title I services, and the dotted lines show the percentile ranks of students identified by teachers as needing Title I services but who were not enrolled in Title I schools and therefore could not receive Title I services.

Figure 3.4 shows that the percentile ranks of needy students in non-compensatory education schools often declined. For instance, while first-grade students receiving Title I reading services in 1976 maintained their average percentile rank of 30, similar students in schools not offering Title I services declined in percentile rank, dropping from 29 in the fall to 25 in the spring. These data suggest that improvements in Chapter 1 students' percentile ranks may actually underestimate the benefit of program participation. For instance, the net benefit of Title I participation for first-grade students was thus actually four percentile ranks, rather than zero.

In nearly all grade levels, the needy students in non-compensatory education schools lost ground in mathematics during the year, whereas the students who received Title I services improved. The patterns were not so clear-cut in reading, however. Comparison students in two grades (four and six) rose in percentile rank over the year, though in each case they started at higher ranks than did Title I students and did not improve as much as the Title I students. Of the 12 comparisons in Figure 3.4 (six

Figure 3.4

Changes in Percentile Ranks for Title I Students and Similar Students Not Receiving Compensatory Education, The Sustaining Effects Study, 1976-77*

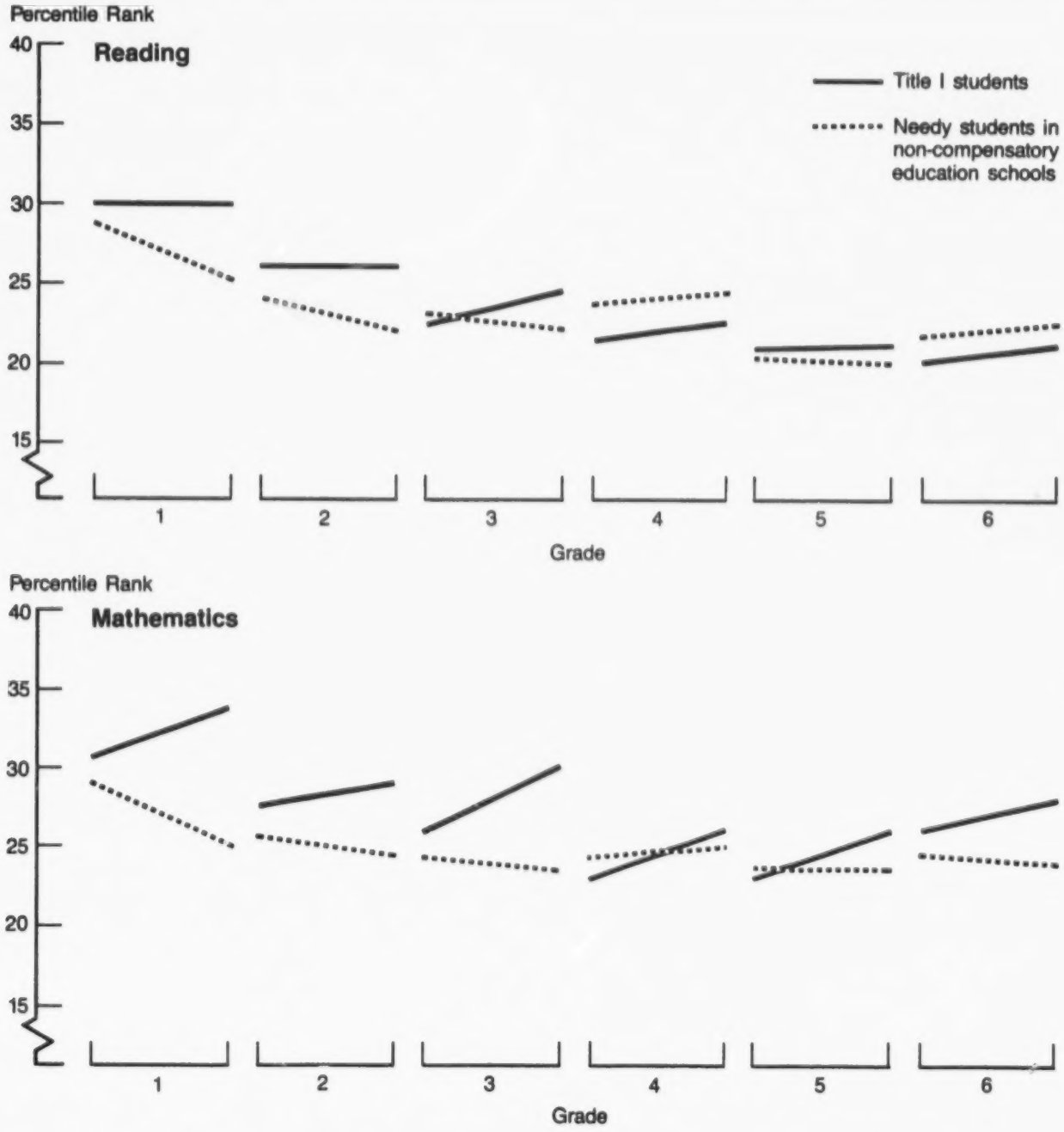


Figure reads: From the fall to the spring testing, 1st grade Title I students receiving reading instruction maintained position at the 30th percentile rank, while needy students in non-Title I schools dropped from the 29th percentile rank to the 25th.

*Percentile ranks presented are based on scores from a fall-spring testing cycle in contrast with the spring-spring cycle used for TIERs data in Figure 3.2. Changes in percentile ranks were calculated by first determining all averages in a standardized scale score metric, and then converting these averages to percentile ranks. See Appendix E for definitions of the measures and Appendix F for the comparable figure presented in NCEs. The number of students included in these analyses varied by grade level and subject; see Appendix G for details.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, *Report 10: Compensatory Services and Educational Development in the School Year.* Santa Monica, CA: System Development Corp., 1981.

grade levels in two subject areas), nine are considered statistically significant.[16] That is, these differences in growth are likely to reflect something more than random fluctuations in test scores. Instead, they reflect real changes.

The Sustaining Effects Study also examined the achievement progress of Title I and needy students in non-compensatory education schools relative to a nationwide representative sample of students. Table 3.1 compares the growth rates of these three groups. Because these comparisons include students who are not comparable to Chapter 1 students, we represent these comparisons in standard deviation units rather than in percentile ranks. Changes in percentile ranks are difficult to compare when students have widely differing achievement levels. At some places in a distribution of scores, a small change in achievement can lead to a large change in percentile rank. Yet in other places, a large change in achievement results only in a small change in percentile rank. This occurs because the largest percentage of students have scores which are concentrated near the average score, while only small proportions of students receive very high or low scores. (See Appendix E for a discussion of these measures.)

Standard deviation units are useful for measuring group differences because the size of a standard deviation unit does not depend on where the students' scores are in the distribution of scores. Standard deviation units are measures of variation among students. These measures are relatively large: the difference between Chapter 1 students and average students is often just slightly more than a single standard deviation unit. If Chapter 1 students begin a school year one standard deviation below average and if the average student gains one standard deviation during the year, Chapter 1 students would need to gain two standard deviations--twice as much as the regular achiever--in order to catch up.

---

[16]The three comparisons not considered statistically significant are those for reading achievement in grades four, five, and six.

Table 3.1 shows that all students gain more in earlier grades than they do in later grades. The representative sample gained more than twice as much in first grade as it did in second grade, relative to the variation among students in each grade (1.98 compared to .87 standard deviation units). Gains among representative sixth graders are about a sixth of their first grade gains (.37 to 1.98 standard deviation units). The differences across grades are not so great in mathematics, but the pattern still is strong. Table 3.1 also shows that needy students who receive no compensatory education gained less than the representative sample in both subjects and in virtually every grade level.

These two groups--the representative students and the needy students--provide two very different norms against which to compare the progress of Title I students. The evidence presented earlier in this chapter suggested that Title I/Chapter 1 students improved their percentile ranks, or relative standing, but that they still performed far below the average student. Table 3.1 shows that these students gained more than comparable needy students on nearly every occasion, but still gained less than the representative sample on half of the comparisons.

Table 3.2 further pursues the differences between Title I gains and the gains of these other two groups by representing the gains of Title I students as a percentage of the gains of these other two groups. It shows, for instance, that Title I students in first-grade reading gained 90 percent of what the representative sample gained, but that they gained 119 percent of what needy students gained when they received no compensatory education. That is, Title I students gained 19 percent more than these other needy students who received no services; but 10 percent less than the representative sample of students. In mathematics the gains of Title I students range from a 10 percent increase in third grade learning rate to a 31 percent increase in sixth grade learning rate, over the learning rates of comparable needy students. Yet

Table 3.1

Growth of Three Groups of Students Participating in the Sustaining Effects Study, 1976-77 (Expressed in Standard Deviation Units[1])

	Representative Sample[2]	Title I Students[3]	Needy Students With No CE[4]
Reading			
Grade 1	1.98	1.79	1.60
2	.87	.85	.77
3	.61	.64	.53
4	.46	.50	.49
5	.42	.38	.34
6	.37	.37	.37
Math			
Grade 1	1.75	1.76	1.40
2	1.24	1.19	1.04
3	1.21	1.13	1.03
4	.84	.90	.79
5	.70	.68	.55
6	.58	.64	.49

1/ All gains are converted to standard deviation units, using the standard deviation of the Fall scores of the Representative Samples.

2/ Data on representative students taken from Tables 1-2 and 1-3, pages 9 and 10 of Report 10.

3/ Data on Title I students from Table 2-2, page 40, Report 10.

4/ Data on comparable needy students from Table 2-5, page 43, Report 10.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, Report 10: Compensatory Services and Educational Development in the School Year. Santa Monica, CA: System Development Corp., 1981.

Table 3.2

Gains of Title I Students as a Proportion
of Gains of Other Students
(From Table 3.1)

	Title I Student Gains as a Proportion of [1]	
	Representative Sample Gains	Needy Student Gains
Reading		
Grade 1	90%	119%
2	98	110
3	105	121
4	109	102
5	90	112
6	100	100
Math		
Grade 1	101%	126%
2	96	114
3	93	110
4	107	114
5	97	126
6	110	131

1/ Title I student gains shown in Table 3.1 are shown here as proportions of the gains of other groups.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, Report 10: Compensatory Services and Educational Development in the School Year. Santa Monica, CA: System Development Corp., 1981.

they still gained less than the representative sample in three of the six grades examined.

It is also possible to present comparisons such as these graphically, though the graphic representation requires still another scaling device. Using a method of scoring the achievement tests that allows growth across all the grade levels to appear within a single scale, Sustaining Effects Study researchers were able to illustrate the achievement progress of three groups, two of which we have already discussed: Title I students and needy students in non-compensatory schools. In this analysis, however, the Sustaining Effects Study researchers did not use the representative sample for its third group, but instead used students not receiving compensatory education though enrolled in Title I schools.[17] This contrast serves roughly the same purpose as the preceding one, in that this group indicates the progress of students who are not considered to be in need of any special services. Figure 3.5 shows the achievement growth experienced by these three groups of students at each grade level. Within each group, we have superimposed the progress of students of different ages into a single figure, so that it simulates the hypothetical progress of a single group of students moving through the entire elementary school sequence. Students labeled in Figure 3.5 as "Title I students" and as "Needy students in non-compensatory education schools" are the same groups shown in Figure 3.4. The growth patterns look different because this new scale is designed to show students' actual learning over time, whereas Figure 3.4 uses a percentile rank scale in order to show changes in relative standing over time. Thus, the downward lines in Figure 3.4, showing a loss in percentile rank,

---

[17]Non-compensatory students in Title I schools are not the same as those in the representative sample used in Tables 3.1 and 3.2. The representative sample included students in Title schools who were not receiving compensatory services, presumably because they were less in need of services, and students in non-Title I schools who were not identified as needing compensatory services. Figure 3.5 includes only the first of these groups.

Figure 3.5

Reading and Mathematics Achievement of Students Receiving and Not Receiving Compensatory Education, Sustaining Effects Study, 1976-77*

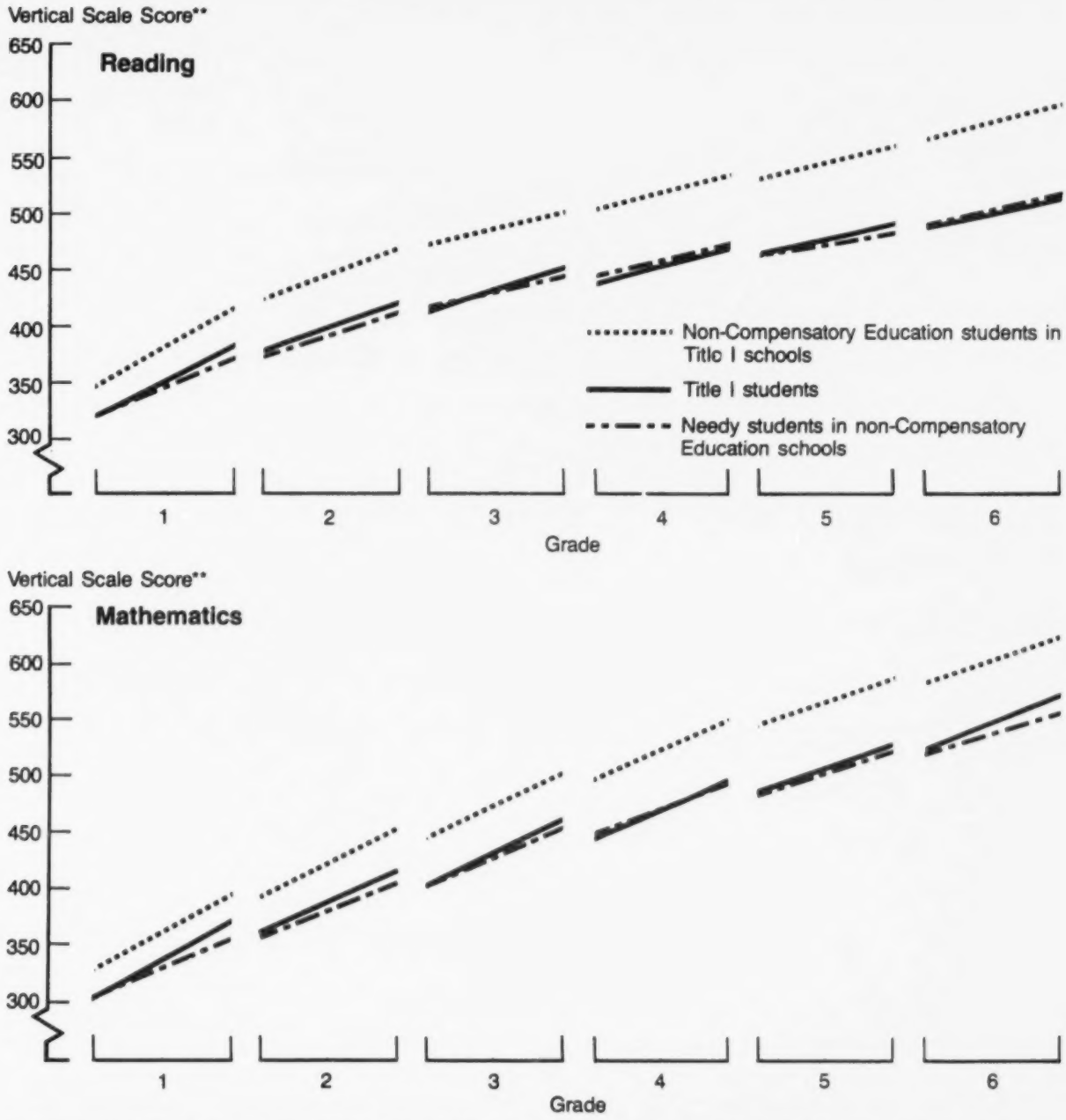


Figure reads: The vertical scale scores of Title I first-grade students for reading and mathematics increased more from the fall to the spring than did those of similar students not enrolled in Title I schools, yet Title I first graders started behind regular first graders in Title I schools who did not receive Chapter I and failed to catch up by the spring.

*The achievement trends presented here differ from those previously used because they are expressed in vertical scale scores rather than percentile ranks. A group can gain in achievement but still show decreases in percentile ranks relative to the entire student population.

**Vertical scale scores are one form of expanded scale scores. They allow comparisons across grade and content levels. See Appendix E. The number of students included in these analyses varied by grade level and subject; see Appendix G for details.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, *Report 10: Compensatory Services and Educational Development in the School Year.* Santa Monica, CA: System Development Corp., 1981.

do not mean that students actually forgot material, but rather that they learned less than other students. Figure 3.5 shows the amount these students actually gained.

Figure 3.5 demonstrates several important trends. First, Title I services reduced the gap between Title I students' achievement and the achievement of more advantaged non-compensatory education students in Title I schools, but only slightly. Second, bearing in mind that these are not the same students followed over six years, but rather six groups of students, each followed during the 1976-77 school year, the data nevertheless indicate that the gap between disadvantaged students and regular students is much larger in later grades than in earlier grades. Although Title I students generally gained more during individual years than comparable needy students in non-compensatory education schools, their rate of growth was not enough during any one year alone to bring them substantially closer to the performance of the more advantaged non-compensatory education students in Title I schools.

Key Assumptions Underlying the Sustaining Effects Study Analyses

The Sustaining Effects Study researchers assessed the progress of Title I students by comparing it both with the progress of comparable "needy" students who received no services and with the progress of non-comparable students who did not need services. The validity of these comparisons depends on two assumptions. First, these comparisons assume that the students in non-Title I schools identified as "needy" were indeed "comparable" to Title I students. Yet we know they attended schools which served relatively fewer poor students than other schools in their districts, and our first report to Congress showed that students attending these schools generally achieved more than students attending schools serving higher concentrations of poor children. Thus the performance of these needy students, even though it often went down relative to national norms, was probably better than we would expect of Chapter 1 students if

they had not received program services.[18]

Second, these comparisons assume that students labeled as "needy" who did not receive Title I services also did not receive any other special educational assistance. If these low-achieving students received extra instructional services because of their low achievement--and there is evidence suggesting that they did (Carter, 1980)--then their achievement gains would reflect that assistance and would overestimate the size of the gain we could expect in the complete absence of compensatory education services. Hence, the benefit of Chapter 1 assistance may be underestimated in our analysis if the comparable "needy students in noncompensatory schools" in Figures 3.4 and 3.5 actually received other forms of educational assistance.

Statistical Corrections for Noncomparable Groups

In recent years new statistical estimation techniques have been developed that more effectively correct for group differences arising from the program's selection procedures. As part of the National Assessment of Chapter 1, we applied these statistical techniques to the Sustaining Effects Study data (Myers, 1986). In this analysis, we formed several different comparison groups whose achievement growth could be used to estimate the expected growth of Title I participants if they had not received Title I services. One of these consisted of all non-participating students in Chapter 1 schools, and another included all non-participating students in non-Chapter 1 schools. The third consisted of a statistically-created group which was comparable to Chapter 1 students both in entering test scores and in family background characteristics. Reanalyses of Sustaining Effects Study data using these analytic

---

[18]The average socioeconomic status of students in a school has been shown to be strongly related to individual student's achievement. Our first interim report to Congress summarized some of this evidence. Other examples of these findings can be found in Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, and York (1966, known as the Equal Educational Opportunity Study--EEOS), McPartland and York (1967), Wolf (1977) and White (1983).

techniques indicated that the more similar the comparison group was to Title I participants, the greater the achievement benefits associated with Title I participation. However, the patterns of effects found when using the statistically optimum comparison group was not substantially different from that found by the original Sustaining Effects Study analysis.

Variability of Outcomes

The findings from these studies should not be taken to mean that effects shown here would be the same for all districts, schools, or students. Many students make larger achievement gains, while others lose ground. Even when scores are averaged across school districts or over entire States, we still see variations in average gains. The State reports submitted to the Department of Education under TIERS illustrate this variability. Table 3.3 presents the range in the average gains reported by States.[19] It shows, for example, that State averages in second-grade reading scores ranged from a loss of 5.7 points to a gain of 7.9 points.[20] Large differences among State average gains are evident in other grade levels as well. Ninth-grade mathematics gains ranged from -2.1 to +15.7 points across States. Further, these variations among States are roughly consistent from grade to grade. That is, States which report small or negative gains in one grade are likely to report similar results in other grades, whereas States which report large gains in one grade are also likely to report large gains in other grades.

---

[19]Table 3.3 includes only those States that reported 500 or more students for a particular grade level. Small sample sizes tend to be affected by extremely low or high scores. By excluding States with small numbers of students in a grade, the results shown in Table 3.3 are more stable and less likely to have been affected by extreme student scores.

[20]Table 3.3 reports achievement gains in the statistical metric used by TIERS, the normal curve equivalent (NCE). Its properties are discussed in Appendix E.

Table 3.3

Range in State-Reported Average Gains* of Chapter 1 Students in Reading and Mathematics in Grades 2 Through 12, 1983-84

	Reading		Mathematics	
Grade	Average gain	Range of State averages	Average gain	Range of State averages
2	1.0	−5.7 to 7.9	3.2	−1.2 to 11.5
3	3.0	−0.5 to 6.0	3.2	−2.5 to 11.4
4	2.9	−0.9 to 9.6	3.1	−2.1 to 7.6
5	3.1	−0.9 to 8.8	4.4	−3.0 to 9.0
6	3.2	−0.8 to 6.9	4.0	−2.6 to 6.7
7	2.5	−1.4 to 6.6	3.5	−1.0 to 6.4
8	2.4	−1.6 to 5.3	3.1	−1.1 to 7.6
9	1.6	−0.4 to 9.5	0.7	−2.1 to 15.7
10	1.1	−1.6 to 6.7	0.5	−1.9 to 2.7
11	0.3	−4.7 to 1.7	1.1	−2.8 to 3.5
12	0.3	−5.5 to 2.8	1.9	0.6 to 4.3

Table reads: From spring 1983 to spring 1984, second grade Chapter I students on average gained 1.0 NCE in their reading scores and 3.2 NCEs in their mathematics scores. However, State gains in average reading scores ranged from −5.7 to 7.9 NCEs while in mathematics they ranged from −1.2 to 11.5 NCEs.

*The average gains presented combine scores for all states submitting reports. Ranges include only those states reporting scores on 500 or more students. Gains are expressed as NCE scores, a metric explained in Appendix E.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology, 1985.

There are a number of reasons why such differences exist. Program effects may vary due to differences in the grade levels or subject matters emphasized, the characteristics and needs of students participating, the extent to which students reside in areas of high concentrations of poverty, the degree to which the achievement test matches the curriculum, the nature and extent of States' oversight of local Chapter 1 programs, the quality of the local Chapter 1 programs, the quality of the regular programs, or the way in which the two programs are coordinated. Many States also sponsor compensatory education programs and these may interact differentially with Chapter 1.

Program Effects on Student Attitudes Toward School

A decade ago, there was considerable interest in student attitudes toward school. Several analysts suggested that attitudes were more important to student achievement in the long run than were short-term achievement gains. This view suggested that a program designed to alter such attitudes could have greater effects on ultimate educational achievement than a program concentrating directly on basic skills. Two studies of Title I--the Instructional Dimensions Study (NIE, 1976; Cooley, 1978) and the Sustaining Effects Study--addressed that hypothesis by measuring student attitudes toward school in addition to student achievement gains. The Instructional Dimensions Study's analyses of students' attitudes towards school revealed no significant changes from fall to spring in either grades one or three, the only two grade levels included in the study. However, the students exhibited very high scores on the attitude instruments in the fall, so that there was little room left for scores to increase by the spring.

Researchers involved in the Sustaining Effects Study found inconsistent patterns in attitudinal changes of participating and nonparticipating students across the six grade levels and two subjects of Chapter 1 instruction. These differences were further

complicated by overall changes in attitude. For instance, all students improved their attitudes in second grade, yet all became more negative during sixth grade.

Some observers have proposed that other indicators, such as attendance records or disciplinary records, be used instead of paper and pencil instruments to measure student attitudes. To date, no national data include such measures.

The Relationship Between Costs and Effects of Chapter 1 Services

In addition to studying the effect of program services on students' standardized achievement test scores, many investigators have also tried to determine the relationship between program effects and the amount of money spent to achieve those effects. But research on this relationship, between program costs on one hand and program effects on the other, has proved to be difficult to conduct for several reasons.

First, estimates of effects by themselves have been difficult to determine, even apart from their relationship to costs. We have already seen that the size of an effect may vary across grade levels, subject matters and types of students, and that the researcher's estimate of the size of the effect depends heavily on the choice of a comparison group against which to compare program beneficiaries.

Second, cost-benefit studies are designed to measure the influence of costs on achievement. Yet decisions about costs--that is, about how to allocate resources--may also be influenced by achievement. Suppose a district allocates its most expensive resources to those students who have the most difficulty learning, while allocating its cheaper resources to eligible students who have less difficulty. And suppose, as a result, all its students gain comparable amounts during the school year. Unless a researcher were able to take these allocation decisions into account, a feat which is difficult to accomplish, the data gathered in this district could indicate no relationship

between costs and effects. For it would show that student growth was uniform, while costs varied considerably.

Finally, districts differ substantially in the nature of their student populations, in their decisions about how to allocate Chapter 1 funds, and in the level of State and local funding available to serve these students. Even though districts may vary by several hundred dollars in their Chapter 1 per-pupil expenditures, they vary by several thousand dollars in their total per-pupil expenditures. In the context of these vast differences in total expenditures, it becomes extremely difficult to isolate the relationship of Chapter 1 expenditures to Chapter 1 effects.

Despite our reservations about research on this topic, some researchers have taken the available evidence as conclusive. Mullins and Summers (1982), for instance review the available evidence and conclude that there is no relationship between compensatory education program costs and benefits. Indeed, a number of investigations have been undertaken to determine the relationship between costs and benefits, and most have been unable to document such a relationship. However, because of the difficult methodological issues involved, our conclusion from reviewing this literature is that researchers have not yet developed adequate methods for determining the relationship between program costs and program effects on standardized achievement test scores.

Summary and Discussion

The analyses described in this chapter lead us to these conclusions:

2. Students receiving Chapter 1 services experience larger increases in their standardized achievement test scores than comparable students who do not. However, their gains do not move them substantially toward the achievement levels of more advantaged students.

The two data sets reviewed here, while differing substantially in date of data collection (1976 versus 1983), identity of data collector (independent researchers versus

school districts), use of comparison groups, and sizes of average gains demonstrated, provide us with remarkably similar patterns of achievement test scores. They both indicate that students participating in local Chapter 1 programs increase their achievement rank, relative to comparable students who received no services. Both also suggest that these gains are not enough to close the gap between Chapter 1 students and their more advantaged counterparts. The two data sets also indicate that Chapter 1 benefits vary from almost none in some subjects and grade levels to relatively substantial benefits in others. Because of the way in which Chapter 1 services are allocated, with students who gain the most removed from the program each year and those who benefit least retained, it is not possible to know whether benefits such as those shown here would add up across the years to yield an aggregate multi-year benefit.

An important difference between the Sustained Effects Study and the TIERS data had to do with the sizes of the average improvements. Whereas the Sustaining Effects Study, conducted during the 1976-77 school year, found elementary reading averages to improve no more than two percentile ranks during the school year, TIERS data collected during the 1983-84 school year show elementary reading score averages improving by two to five percentile ranks, depending on the grade level. And whereas the Sustaining Effects Study showed average mathematics scores generally improving by three or four percentile ranks during a school year, TIERS indicates that such scores improved by five to seven percentile ranks. These differences are particularly remarkable in light of the fact that the Sustaining Effects Study used a fall-to-spring testing schedule, a schedule now generally believed to inflate growth estimates (Keesling, 1981).

Several hypotheses could be generated to account for these differences, but two are particularly important. One is that program administrators and teachers have

actually improved their practices in the eight years between these two analyses, and that the TIERS data reflect the effects of these improvements. The second hypothesis is that, because TIERS permits local districts to choose their own tests, whereas the Sustaining Effects Study used a common test across all districts and schools, the TIERS data more accurately reflect the instructional objectives of local Chapter 1 projects and consequently are able to document program effects that a common test might not measure. Freeman and others (1983) have shown that the extent to which test content matches the content of textbooks can vary substantially from one combination of textbook and test to another. Each of these hypotheses probably has some merit, but the relative importance of each--or of others--cannot be ascertained from the available data.

3. Students participating in Chapter 1 mathematics programs gain more than those participating in Chapter 1 reading programs

Findings from both TIERS and the Sustaining Effects Study indicate that mathematics programs yield larger gains in student achievement than do reading programs. This was true for grades two through twelve in the TIERS data and for grades one through six in the Sustaining Effects Study. When the Sustaining Effects Study researchers applied statistical tests to their data, they found that reading programs yielded statistically significant gains only in grades one, two, and three whereas mathematics programs yielded significant gains in all six grades.

Several hypotheses have been put forward to account for this finding. One hypothesis is that mathematics is provided to students who have higher entering achievement scores than students in reading programs and that such students are more likely to benefit from compensatory education services. Another is that reading and mathematics differ in the extent to which they are taught outside of Chapter 1. Because reading is a task required by many school subjects, the extra instruction provided by Chapter 1 may constitute a smaller portion of the student's total reading

practice during a school day. Mathematics, on the other hand, is not used in as many other school subjects and thus the additional mathematics instruction provided by Chapter 1 programs may constitute a significant addition to the students' total exposure to mathematics.

4. Students in early elementary Chapter 1 programs gain more than students participating in later-grade programs.

The TIERS data show a consistent pattern of decreasing gains as students moved from elementary to middle to secondary grades, with program effects almost negligible in the secondary grades. The earlier Sustaining Effects Study yielded a more mixed pattern within the elementary grades.

Why such differences occur is not clear. Gains in upper grades could be smaller because students start further behind in these grades, and generally have more difficulty learning; because they are more disaffected with school and are less motivated to learn; because the academic content taught in these grades is more difficult to learn; because the services provided are not as good as those provided to younger students; or because secondary-level achievement tests do not cover the types of basic skills taught by secondary Chapter 1 programs.

5. Evidence regarding program effects on student attitudes toward school is inconclusive

Though two studies of Title I attempted to measure changes in student attitude toward school, and to attribute these changes to Chapter 1, neither was successful. Part of the difficulty is in measuring attitudes by means of paper-and-pencil instruments. Some analysts now advocate the use of such indicators as attendance rates or participation in extracurricular activities to estimate attitudes toward school.

6. Researchers have not yet developed adequate methods for determining the relationship between program costs and program effects on standardized achievement test scores.

Though a few studies have attempted to relate program costs to program benefits, the task has proved to be extremely complicated and no studies to date have been able to solve the many methodological problems presented.

4. Sustaining Achievement through the Summer and Future School Years

Central Findings

7. The achievement gap between disadvantaged and advantaged students appears to widen during the summer months.

8. Title I-supported summer programs have not narrowed the achievement gap; however, these programs often were not designed to be academically rigorous.

9. Students who discontinue Title I appear gradually to lose ground when they no longer receive services.

10. Chapter 1 students with very low achievement levels appear to maintain their relative achievement levels while participating in Chapter 1, but not to move ahead. However, the evidence suggests they would have lost ground relative to their peers if they had not received services.

11. No nationally-representative studies have examined the long-term effect of Chapter 1 programs on graduation rates, future education, or adult literacy.

Patterns of achievement test scores among disadvantaged students suggest that these students not only score lower, on average, than their more advantaged peers, but that their scores fall farther and farther behind their more advantaged peers as they move through school. Why this happens is not understood and has been the subject of both speculation and research, some of which will be described in this chapter. One hypothesis proposed by Hayes and Grether (1969) received considerable attention in the 1970s: disadvantaged students learn as much as other students during the school year, but they forget more or learn at a lower rate during the summer. Such a phenomenon would be consistent with one of the underlying premises of the Federal compensatory education program: students intended to receive Chapter 1 services were assumed to come from homes that could not provide the enriching experiences that more affluent homes provided. It was assumed, for example, that the homes of disadvantaged students would offer fewer reading materials and fewer opportunities to engage in learning activities. If this is so, then disadvantaged students would lose ground, relative to their more advantaged peers, each summer.

The possibility of a summer "dropoff" in the achievement of disadvantaged students raised three questions. First, was it true? Second, if it was true, would summer school programs supported by Chapter 1 prevent such losses? And third, did the achievement gains that occurred as a result of Chapter 1 services during the regular school year last through the summer, or were they also lost? Each of these questions has received research attention.

Extent of a Summer Dropoff in Achievement

With regard to the first question--is there a summer dropoff in the achievement of disadvantaged students--two studies of Title I examined students' test scores over time: the NIE Compensatory Education Study (NIE, 1978) and the Sustaining Effects Study (Klibanoff and Haggart, 1981). Both found that the achievement of disadvantaged students' did not in fact decrease during the summer. Instead, it increased. However, it did not increase as much as did the achievement of advantaged students. Thus, the notion of summer loss did not account for the increasing gap in achievement as students progressed through school, but differences among childrens' rates of increase over the summer did.

Table 4.1 shows the findings from the NIE Compensatory Education study. The table presents both school-year and summer gains in the achievement scores of compensatory education students who began the school year below the 50th percentile, and noncompensatory education students who began the school year scoring above the 50th percentile. These data indicate that all students gained more during the school year[21] than during the summer, but that most groups gained during the summer as well. It also indicates that summer gains were larger in reading than in mathematics.

---

[21]The gains of Chapter 1 students shown here are not representative of nationwide averages. Because the NIE wanted to study the relationship between project characteristics and student outcomes, it purposely selected well implemented, stable projects for its study.

Table 4.1

Gains in Achievement for Compensatory and Non-Compensatory Education Students* During the School Year and Summer, 1976-77

Achievement Gains [Expanded scale scores]**

			School year gain	Summer gain
Compensatory Education Students	Reading	Grade 1	69	0
		Grade 3	44	8
	Mathematics	Grade 1	43	2
		Grade 3	64	−1
Non-Compensatory Education Students	Reading	Grade 1	56	10
		Grade 3	36	21
	Mathematics	Grade 1	39	6
		Grade 3	62	7

Table reads: Compensatory education students who were in 1st grade and receiving reading services gained 69 units in reading over the school year but 0 units over the summer. In comparison, 1st grade pupils not receiving compensatory reading instruction gained 56 units in their reading score during the school year and 10 units during the summer.

*Students in compensatory education included only those whose pretest scores were below the 50th percentile; those not receiving compensatory education included only those whose pretest scores were at or above the 50th percentile.

**Gains are expressed as a form of expanded standard scores; these are defined in Appendix E.

Source: National Institute of Education, *Compensatory Education Study. Final Report to Congress From the National Institute of Education.* Washington, D.C. September 1978.

Most important for Chapter 1, however, is the difference in summer gains between the compensatory education students and other students. In both subjects and at both grade levels, noncompensatory education students gained more during the summer than did compensatory education students.

These patterns were corroborated by Heyns (1978) in one of the most influential studies to date on this issue. Heyns analyzed data from one large urban school system and concluded that schooling mitigates the increasing achievement gap among students from diverse backgrounds, and that the summer months increase the achievement gap. Table 4.2 presents a portion of Heyns' data. Heyns' data do not identify Chapter 1 students specifically, but they compare the summer gains of low income students and higher-income students. Students from families with low incomes displayed substantially smaller summer gains than did students from families with higher incomes. This difference is especially pronounced among black students: black sixth graders from low-income families demonstrated achievement score losses over the summer in word knowledge, while black students from higher-income families showed gains. To the extent that these summer losses among low-income blacks, and smaller-than-average gains among low-income white students, are cumulative they help explain the growing achievement gap between disadvantaged and other students.

In a more recent paper, Heyns (1986) used the Sustaining Effects data to illustrate the widening gap between low and high-achieving students. Figures 4.1 and 4.2 show the achievement growth patterns for students in the lowest, second, and third quartiles. In each figure, the school year and summer growth of students at all grade levels are superimposed to simulate the growth of a single group across six grades. The scores are presented in vertical scale scores, a metric that permits student achievement in all six grade levels to appear on a single scale. Both charts suggest that groups who enter first grade with different achievement levels can be expected to

Table 4.2

School Year and Summer Achievement Gains by Race and Family Income for Sixth Grade Pupils in One Large Urban District, 1971-72 (Grade Equivalent Scores)*

Race and income	Fall 1971	Spring 1972	Fall 1972	School gains	Summer gain	N
National average**	5.10	5.80	6.10	.70	.30	
Total sample***	4.25	4.87	4.86	.62	−.01	1,445
White	4.96	5.80	6.04	.84	.24	459
Less than $9,000	4.21	4.86	4.93	.65	.07	109
$9,000 - $14,999	4.77	5.73	5.91	.96	.18	115
$15,000 +	5.86	6.86	7.15	1.00	.29	124
Black	3.93	4.44	4.32	.51	−.12	986
Less than $4,000	3.62	4.04	3.76	.42	−.28	187
$4,000 - $8,999	3.84	4.35	4.23	.51	−.12	325
$9,000 - $14,999	4.08	4.67	4.55	.59	−.12	169
$15,000 +	4.57	5.19	5.41	.62	.22	101

Table reads: From fall 1971 to fall 1972, the performance of 6th grade white students in the sample increased .84 grade equivalent units during the school year and .24 units during the summer.

*Achievement is expressed as students' mean grade equivalent scores on the Word Knowledge subtest of the Metropolitan Achievement Test; grade equivalent scores are described in Appendix E.

**These are the mean grade equivalent scores predicted for the particular test dates.

***The sample consists of all students with test scores available at all three dates. Totals include students for whom family income was missing.

Source: B. Heyns, *Summer Learning and the Effects of Schooling*. New York: Academic Press, 1978, pp. 45, 282-283.

Figure 4.1

Vertical Scale Scores for Reading as a Function of Grade Level for Selected Percentile Ranks*

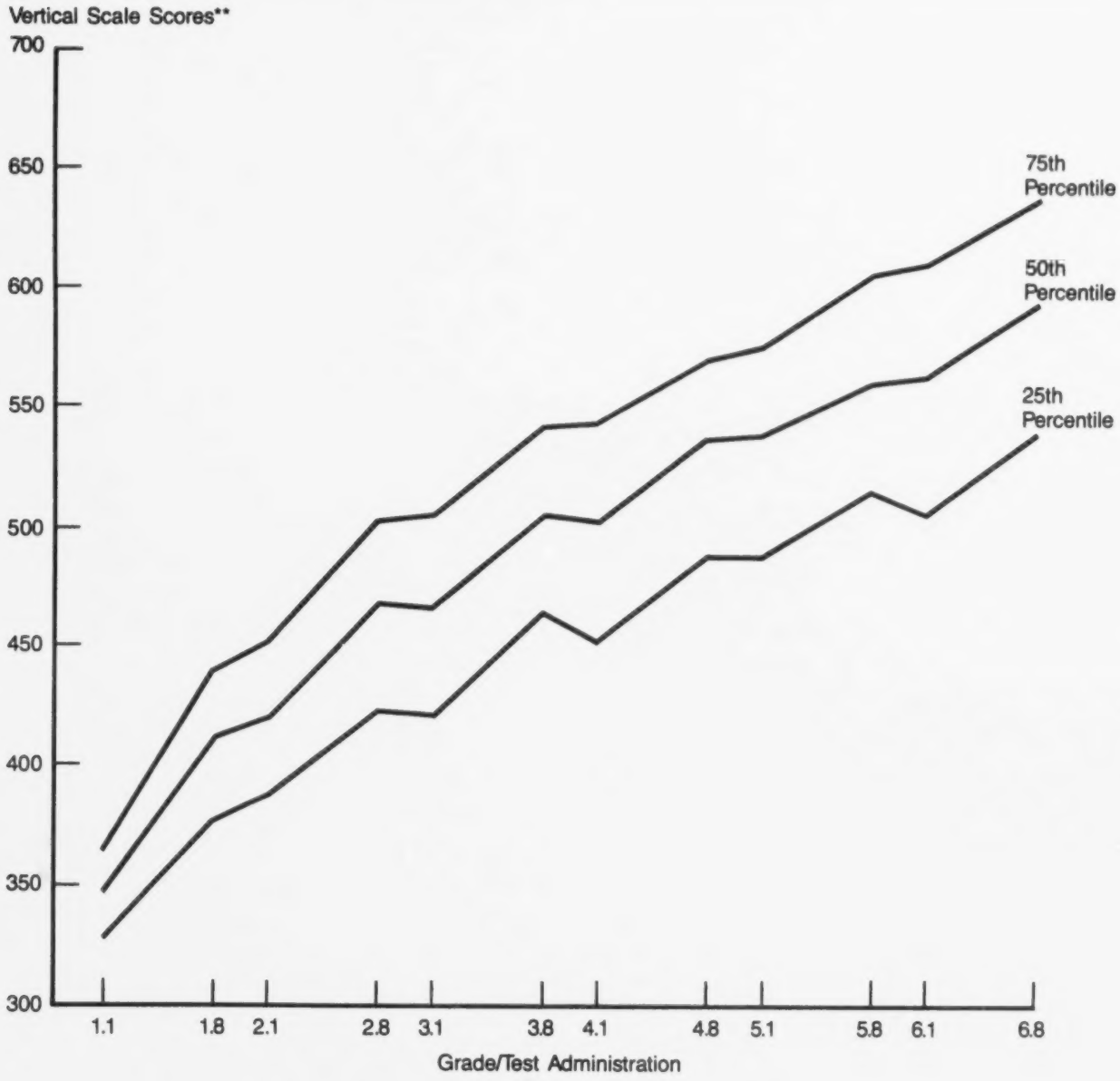


Figure reads: A beginning first grade student scoring at the 25th percentile, which corresponds to a vertical scale score (VSS) of 326, typically received a VSS the next spring of 375 and a VSS of 386 the following fall as a second grader.

*The zig-zag nature of the curves should not be attributed necessarily to "summer drop-off". The "negative growth" observed for some spring to fall intervals may be due to sample and test-level differences.

**Vertical Scale Scores are expanded scale scores as defined in Appendix E.

Source: J. Hemenway, M. Wang, C. Kenoyer, R. Hoepfner, M. Bear, and G. Smith. *Report 9: The Measures and Variables in the Sustaining Effects Study.* Santa Monica, CA: System Development Corp., 1981.

Figure 4.2

Vertical Scale Scores for Mathematics as a Function of Grade Level for Selected Percentile Ranks*

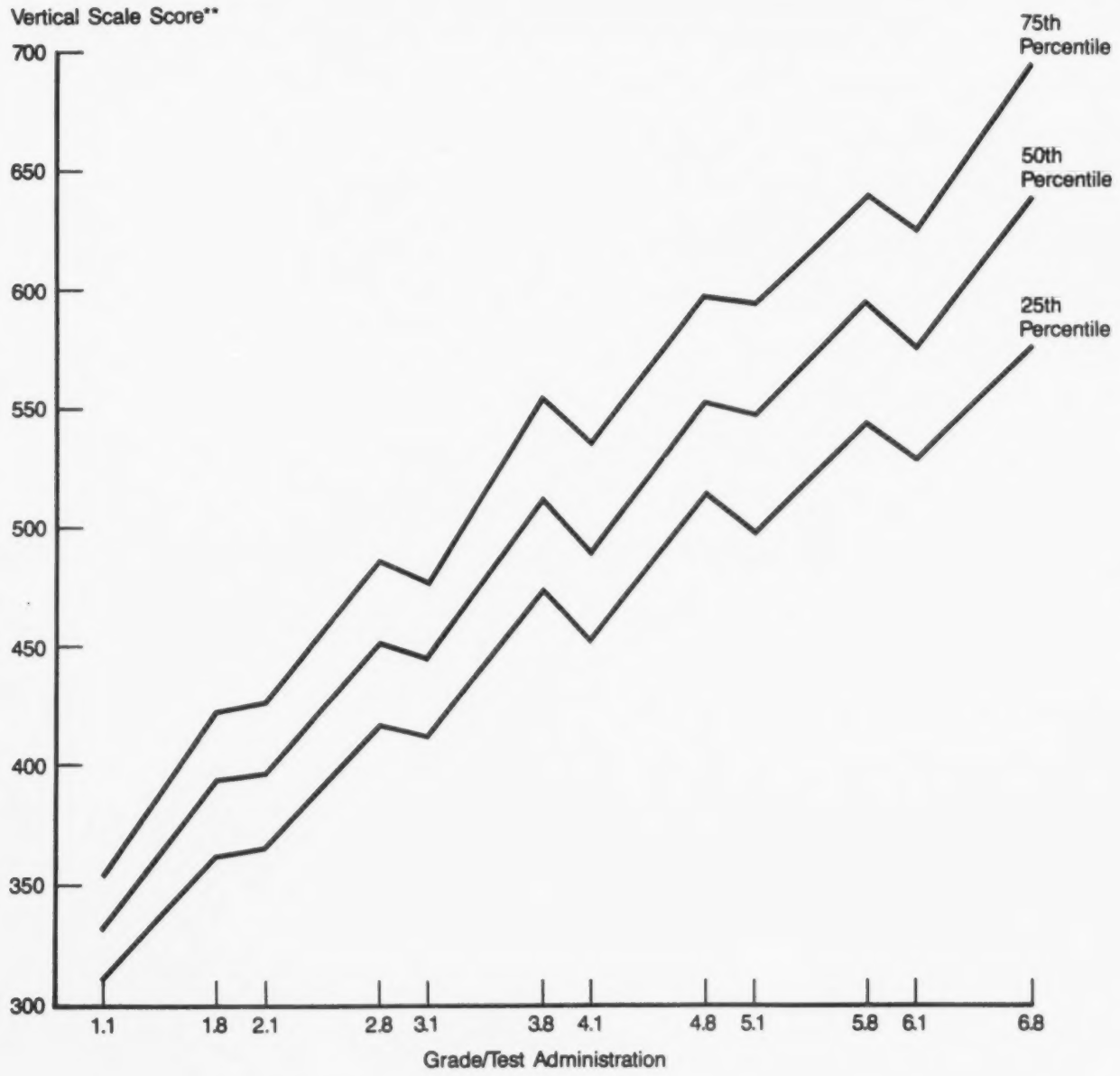


Figure reads: A beginning first grade student scoring at the 25th percentile, which corresponds to a vertical scale score (VSS) of 310, typically received a VSS the next spring of 362 and a VSS of 366 the following fall as a second grader.

*The zig-zag nature of the curves should not be attributed necessarily to "summer drop-off". The "negative growth" observed for some spring to fall intervals may be due to sample and test-level differences.

**Vertical Scale Scores are expanded scale scores as defined in Appendix E.

Source: J. Hemenway, M. Wang, C. Kenoyer, R. Hoepfner, M. Bear, and G. Smith. *Report 9: The Measures and Variables in the Sustaining Effects Study.* Santa Monica, CA: System Development Corp., 1981.

grow farther apart as they move through the elementary grades. The later-grade differences are even greater in reading than they are in mathematics. Because Figures 4.1 and 4.2 superimpose the achievement growth of separate groups of students, one at each grade level, to create the appearance of a single group moving across the grades, the summer "growth" is difficult to interpret. Both figures indicate, though, that students beginning school in the fall often have scores that are lower than those of students completing the preceding grade. And, in reading at least, the summer dropoffs are more severe among low-achieving students than among high-achieving students.

These increasing disparities among achievement scores are particularly noteworthy when viewed together with the amount students learn each year. During first grade, for instance, students at all achievement levels gain a substantial amount. Yet the gap is already widening, and by the end of first grade students in the lowest group have scores comparable to those held by the highest group at the beginning of first grade. By the time students reach sixth grade, their achievement scores are much more diverse. Yet, as we saw earlier, the amount students gain during the sixth grade is much less, relative to the differences among high- and low-achieving students, than it is in earlier grades, especially in reading. Thus students in the lowest group begin the sixth grade with reading scores comparable to those held by the highest group when they began third grade, and do not gain nearly enough during the year to close the achievement gap. By the end of sixth grade, their achievement is still comparable to the third-grade reading achievement of the highest group.

Impact of Summer School Programs

The notion that summertime may contribute to the widening achievement gap sparked interest in the 1970s in the hypothesis that Title I summer projects might either prevent summer losses or promote further growth in the achievement scores of

educationally-disadvantaged students. Both the Sustaining Effects Study researchers and Heyns examined the effects of participation in summer programs.

The Sustaining Effects Study identified students who participated in Chapter 1 summer programs and compared their achievement growth with that of comparable students not participating in summer programs. Figure 4.3 shows the resulting achievement growth patterns. It suggests two important trends. First, compared with students who did not attend summer school, students who attended summer school had lower achievement scores initially and thus were more in need of compensatory education services. Second, summer school programs had very little apparent effect on these students' achievement. In most cases, the achievement growth of summer school students was still less during the summer than it was during the school year, and it was rarely greater than the summer growth of the higher-scoring non-participating students. The Sustaining Effects Study analysts attributed this finding to the fact that Title I summer programs were typically brief, averaging about six weeks, and that the summer instruction often was neither intense nor geared specifically towards increasing basic skills (Klibanoff and Haggart, 1981). Thus, the programs could not realistically be expected to affect achievement test scores. Heyns' research also indicated that summer programs are generally not rigorous in an academic sense and thus cannot be expected to raise achievement test scores (Heyns, 1978).

Retention of Chapter 1 Benefits Obtained
During the School Year

The third important question regarding the relationship between summer and school-year learning was this: since the achievement gap between disadvantaged and other students seems to widen over time, to what extent do the benefits of school year Chapter 1 services remain through the summer and into the following school year?

Figure 4.3

Gains in Achievement for Title I Students Attending and Not Attending Summer School, Summer 1977

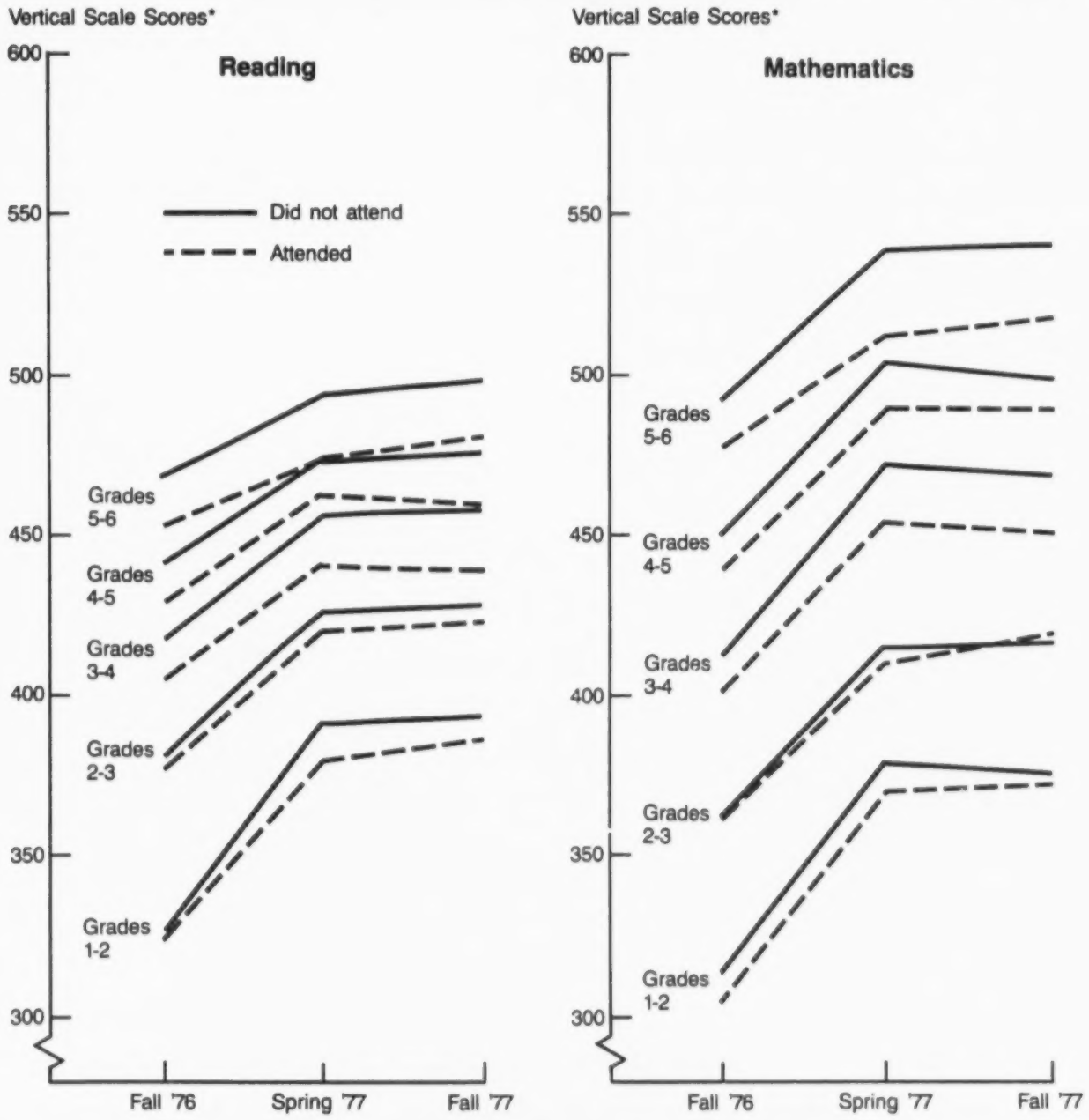


Figure reads: Title I students who were in grade 1 during fall '76 who did not attend summer school had slightly higher test scores at all three testing periods (fall '76, spring '77, and fall '77) than did students who attended summer projects.

*Vertical scale scores are a form of expanded scale scores as defined in Appendix E. The number of students in these analyses varied by grade level and subject; see Appendix G for details.

Source: L. Klibanoff and S. Haggart. *Report 8: Summer Growth and the Effectiveness of Summer School.* Santa Monica, CA: System Development Corp., 1981.

TIERS provides one source of evidence on this topic. Under TIERS, school districts can choose from several options in deciding when to test students. They may test students once each year, every spring or every fall, or they may test students twice a year, in the fall and again in the spring. When the former schedule is used, a single test serves both as a posttest for one school year and as a pretest for the next school year. Further, the period between pretests and posttests is a full year, including summer, whereas the period between fall and spring testing includes only the school year itself.

The magnitude of students' apparent achievement gains varies considerably depending on which of these testing schedules districts use (Keesling, 1981). Districts using annual testing cycles report achievement gains that are smaller than those of districts using the fall-spring schedule. To illustrate the difference, Figure 4.4 superimposes test results from these two schedules. The figure indicates that second graders tested on a fall-spring schedule, for example, started their reading programs scoring at the 21st percentile rank and finished with scores at the 36th percentile rank, for an apparent gain of 15 percentile ranks (the dashed line in the figure). Yet those tested on an annual schedule (the solid line) moved from the 29th percentile rank to the 31st, for an apparent gain of only two percentile ranks.

This figure indicates two important points. First, the amount of improvement observed in these students' achievement depends on the school districts' testing schedule. In both reading and mathematics, program assessments based on annual testing schedules yield much smaller achievement test score gains than those derived from fall-spring testing schedules. Second, many differences, especially in mathematics, appear to be due more to differences in <u>starting</u> achievement levels than

Figure 4.4

Changes in Percentile Ranks for Reading and Mathematics by Testing Cycle, 1983-84*

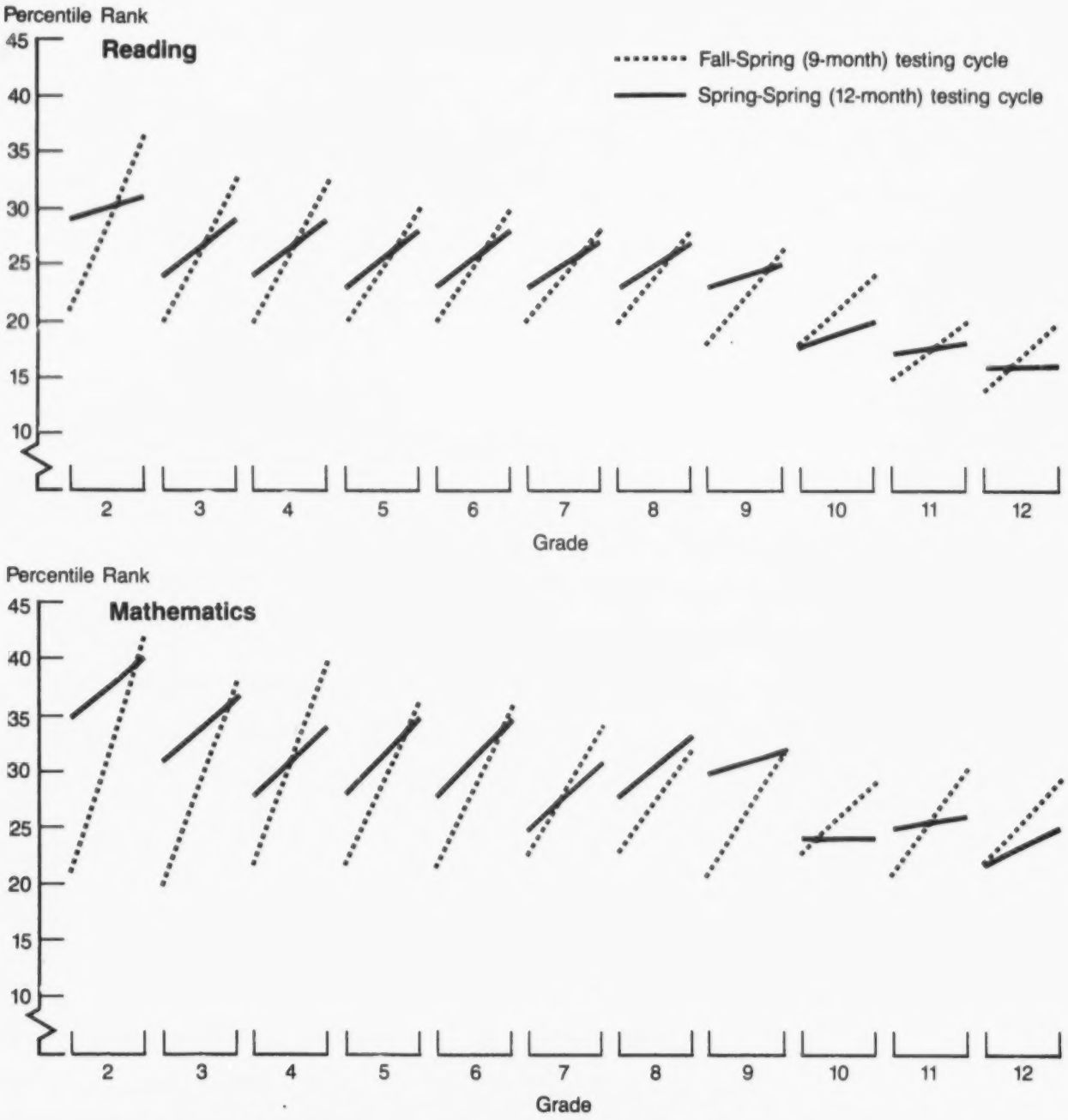


Figure reads: The percentile rank of the average second grader receiving Chapter I reading instruction increased from the 21st to the 36th using a fall-spring testing cycle, while a spring-spring testing cycle resulted in a change from the 29th to the 31st percentile rank for similar students.

*Changes in percentile ranks were calculated by first determining all averages in normal curve equivalents (NCEs), a standardized scale score metric, and then converting those averages to percentile ranks. See Appendix E for definitions of the measures and Appendix F for the comparable figure presented in NCEs. The number of students included in these analyses varied by grade level and subject; see Appendix G for details.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology. 1985.

in ending achievement levels.[22] That is, students tested in the fall demonstrate much lower entering achievement scores than those tested the preceding spring, even though, in several cases, they demonstrate comparable scores at the end of the school year.[23] It is possible, of course, that these disparities reflect differences in the characteristics of the students enrolled or in the effectiveness of programs in districts employing these two testing schedules. But several other important possibilities have been suggested to account for these patterns.

One hypothesis, presented by Keesling (1981), has to do with local scheduling of fall test administrations relative to the schedules recommended by test publishers. Student achievement improves rapidly during the first several weeks of the school year, presumably as students get back into the spirit of the academic enterprise. It is possible that, if districts tested students in the sixth week of the school year, rather than the first, for instance, much of the inflated portion of the fall-to-spring gain would be removed. Without more knowledge of the exact dates when districts administer their fall tests, it is not possible to know how much of the fall to spring gains are due to this early school-year phenomenon. However, we do know that the Chapter 1 Technical Assistance Centers work with districts to improve their testing schedules.

---

[22]Keesling (1981) presented 1979-80 TIERS data showing more comparable post-test data then we find in the 1983-84 data.

[23]Despite the remarkable differences in outcomes between testing schedules, the outcomes suggested by the additional testing schedule shown in Figure 4.4 reinforce the general patterns described in Chapter 3. Both testing cycles shown in Figure 4.4, for instance, indicate that older students enter reading programs at lower percentile ranks than younger students and that older students gain relatively less during the school year. Both estimates indicate higher relative starting ranks of students in mathematics programs, and both suggest that mathematics gains are greater among younger students than among older students.

Another hypothesis, also discussed by Keesling (1981), is that districts may inadvertently inflate fall-to-spring gains by the way they administer their tests. They may, for instance, be less careful in the fall because they don't perceive the fall test as important. Then, since the spring test measures student growth during the school year, teachers may work hard to prepare students for the spring test and to motivate students to do their best. In addition, Linn and others (1982) have listed a number of technicalities which could contribute to an artificial inflation of the fall-to-spring gains.

Finally, there is the hypothesis that disadvantaged students gain less during the summer than other students do, and consequently lose their relative rank. If these summer losses were entirely recovered in the early fall of each school year, as Keesling and others seem to suggest, then the patterns shown in Figure 4.4 reflect nothing more than an artifact of testing practices and momentary but inconsequential student forgetting. If, on the other hand, these summer experiences accumulate over time, contributing to a continual widening of the achievement gap between advantaged and disadvantaged students, then the patterns shown in Figure 4.4 reflect an important phenomenon regarding the learning patterns of disadvantaged children.

Probably each of these phenomena--technical details of test administration and students actually gaining relatively less in the summer--contribute to this pattern of test data. Whatever the reasons, the remarkable differences between the apparent achievement gains indicated by these two testing schedules does suggest that at least some portion of the Chapter 1 benefit is lost during the summer.

Sustaining Achievement Gains Across School Years

Concerns about the learning rates of disadvantaged students prompt two separate questions. First, what are the effects of continued services for those students who remain in the program for two or three years? Second, are the gains demonstrated

while students participate in Chapter 1 maintained after students no longer receive services? Each of these questions will be addressed in this section.

Figures 4.5 and 4.6 present achievement scores of students who participated in Title I reading (Figure 4.5) and mathematics (Figure 4.6) programs for different numbers of years during which the Sustaining Effects Study collected data. The lowest section in each figure shows the achievement profile of students who received Title I services all three years. In the upper two sections of Figures 4.5 and 4.6, achievement patterns are shown for students who participated in Title I for one or two years, respectively, of the three possible years. In these upper sections, separate lines are presented for students who participated in Title I for different subsets of years. For instance, the top section displays separate growth patterns for students who participated during the first year only, the second year only, or the third year only. The middle section portrays students who participated during years one and two, years two and three, and years one and three. For each group of students, the solid portion of the line indicates that the group participated in Chapter 1 at that time, while the dashed portion indicates non-participation. The achievement scores shown in Figures 4.5 and 4.6 were recalibrated each year. Under this method, the average achievement score is always 100, and changes in scores are measured relative to peers, just as changes in percentile ranks are. Thus, a strictly horizontal line indicates that achievement has improved at an average rate; there has been no relative gain or loss.

Figures 4.5 and 4.6 indicate that participation is related to starting achievement. Students participating for only one of the years, regardless of which year it was, started with higher achievement scores than students who participated for either two or three years. Conversely, students participating for all three years tended to start at lower achievement levels, relative to their peers, (starting scores in reading were 81, compared to 85 for two-year participants in reading programs and over 90 for one-

Figure 4.5

Reading Achievement of Students with Different Patterns of Participation in Title I Across Three Years, The Sustaining Effects Study, 1976-79

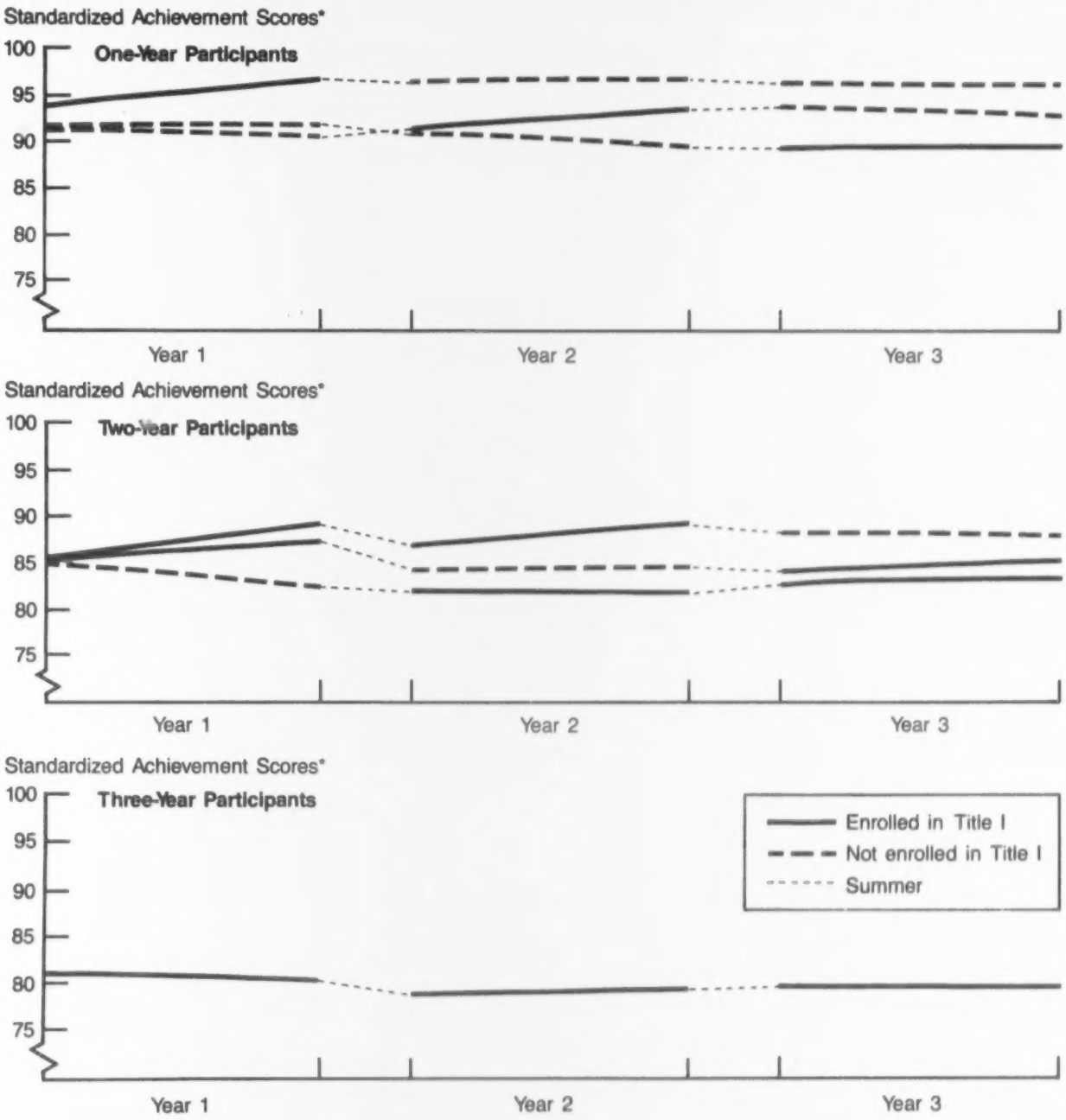


Figure reads: Students who participated in Title I only during the first year of the study began with an average score of 94 and ended their first year with an average score of 97. Their scores declined slightly across the next two years. They completed the third year with an average score of 95.

*These are a form of standard scores, as described in Appendix E, which have been standardized so that the mean at each test period equals 100 and the standard deviation equals 20. The number of students in these analyses varied by grade level and pattern of participation; see Appendix G for details.

Source: L. Carter, *A Study of Compensatory and Elementary Education: The Sustaining Effects Study, Final Report.* Santa Monica, CA: System Development Corporation, January 1983.

Figure 4.6

Mathematics Achievement of Students with Different Patterns of Participation in Title I Across Three Years, The Sustaining Effects Study, 1976-79

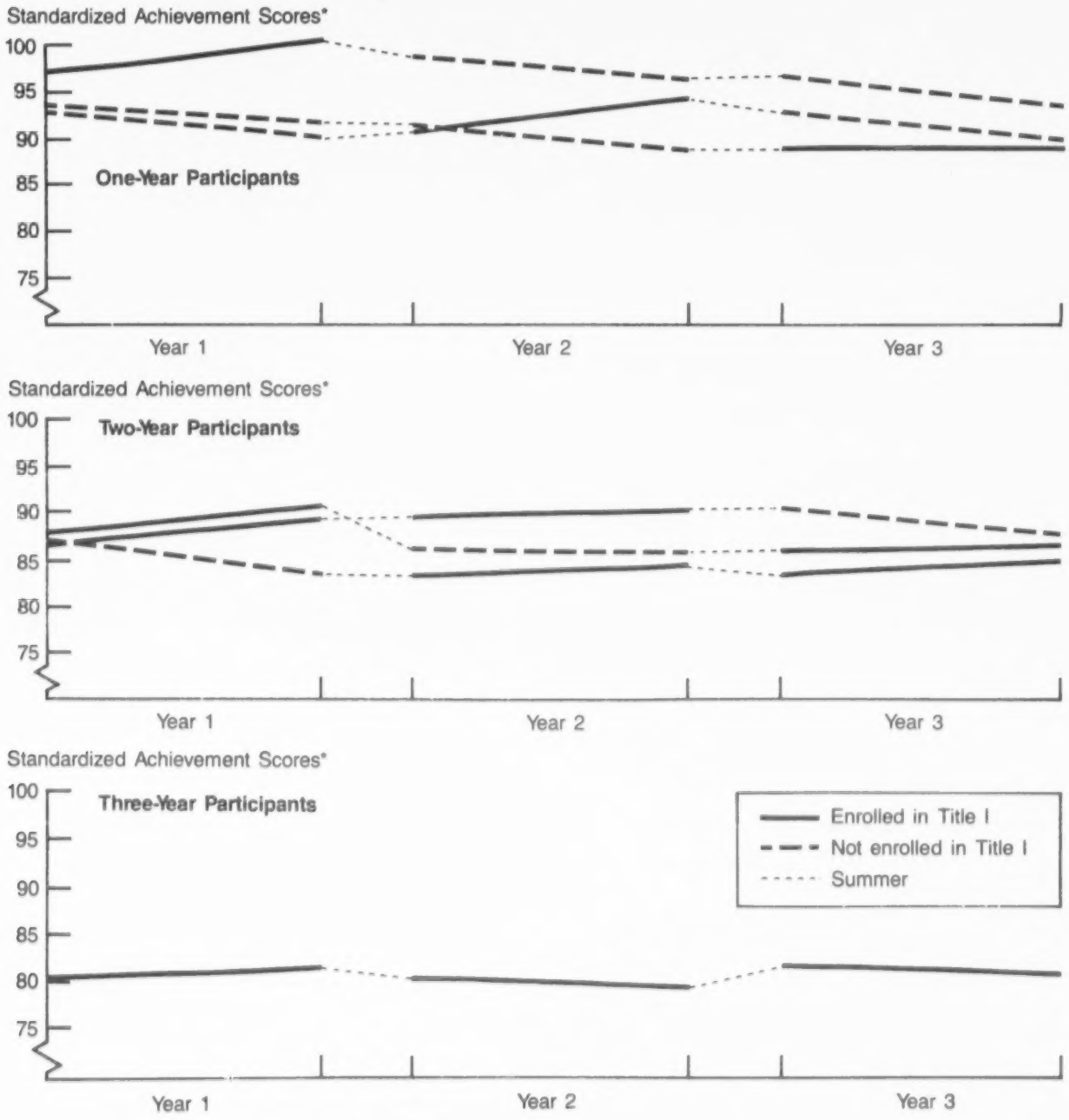


Figure reads: Students who participated in Title I during the first year only started with an average score of 97 and finished the year with an average of 101. Thereafter their scores declined and they completed the three-year period with an average score of 93.

*These are a form of standard scores, as described in Appendix E, which have been standardized so that the mean at each test period equals 100 and the standard deviation equals 20. The number of students in these analyses varied by grade level and pattern of participation; see Appendix G for details.

Source: L. Carter, *A Study of Compensatory and Elementary Education: The Sustaining Effects Study, Final Report.* Santa Monica, CA: System Development Corporation, January 1983.

year participants in reading programs) and to remain at those low levels throughout the entire three-year period of data collection. Figures 4.5 and 4.6 also show that achievement scores rise while students are participating in Title I and remain stable or decline when students do not participate.

With regard to our first question on the merits of multi-year participation relative to one-year participation, these figures indicate that three-year participants, who began with the lowest test scores in both reading and in mathematics, appear to have done no more than maintain their relative positions throughout these three years. However, given earlier evidence that comparable low-scoring students actually lose their relative standing without special services, these horizontal patterns of test scores may indicate that Chapter 1 services have actually improved these students' achievement.[24] The two-year participation data also do not yield a clear pattern with regard to accumulating benefits. For students who participated in reading programs during year one and two of the study, for instance, the second year appears not to add appreciably to what was accomplished during the first year, in part because some of the first year benefits were lost during the summer. Those who participated in years two and three demonstrate a similarly ambiguous pattern this time because of summer time improvements. In contrast, students who participated in mathematics programs during years one and two appear to have benefitted just as much from each year, and to have experienced no summer loss in between. For them, the benefits of the second year are clearly over and above the benefits of the first year. Those who participated in mathematics programs during years two and three demonstrate a similar accumulation of benefit. The Sustaining Effects researchers tried to test the effects of multi-year

---

[24]Figures 4.5 and 4.6 actually depict test-score performance rather than percentile ranks. However, these tests scores were re-calibrated on each administration, so that the average is always 100. Consequently, the growth patterns can be interpreted analogously to those indicated by percentile ranks. That is, a horizontal line does not mean "no growth". It means "no change in relative standing".

participation for different grade level and subject matter combinations, but found a similar confusing picture (Wang, et al, 1981).

With regard to our second question, about the extent to which gains made while participating in Chapter 1 are retained once students leave the program, we rely on students who received service only during the first one or two years of the Sustaining Effects Study so that we can observe their achievement patterns during the school years following program participation. The one-year sections of Figures 4.5 and 4.6 show that students who received services only during the first year of the Sustaining Effects Study gained in both subject areas while participating, but lost ground relative to their peers in later years. Students who had received first-year assistance in mathematics appear to have lost more in later years than those who received first-year assistance in reading. Because these students entered the study with higher scores to begin with, however, their achievement declines still did not leave them with scores as low as those of two-year or three-year participants.

The middle sections of Figures 4.5 and 4.6 show that students who participate during the first two years, but not during the third, also gained during their participation, regardless of subject area. Those who received reading assistance during the first two years appeared to maintain their position during the third year, while those who received mathematics assistance during the first two years declined slightly during the third year. The middle portions of Figures 4.5 and 4.6 also indicate the progress of students who received services during the first and third years of the Sustaining Effects Study. These students appear to have benefited substantially from their first year of service, but they also experienced substantial summer losses

following their first year of participation.[25] Their scores were maintained during the second year, but they were again selected to participate in Title I during the third year.

Generally speaking, these patterns suggest that students gain more than expected while they participate in Chapter 1, and in mathematics these benefits appear to accumulate across years. On the other hand, again in mathematics, students gain less than expected when they do not participate. The gains of previous years do not appear to help students accommodate future learning demands. Most former Chapter 1 students maintained their relative standing in reading, but former Chapter 1 mathematics students gained much more slowly without program assistance, and consequently lost more ground relative to their peers.

In a related substudy, the Sustaining Effects Study investigators also followed some of these Title I students over succeeding school years until they reached junior high school. They found that the earlier achievement increases from Title I services had virtually disappeared by the time the students reached junior high school and that former Title I students enrolled in more remedial courses in junior high school than did comparable students. Thus, although elementary school Chapter 1 services positively influenced student achievement, they did not improve achievement enough to eliminate students' need for further special assistance (Carter, 1984).

A more recent study, undertaken by personnel of the Chapter 1 Technical Assistance Centers, under the direction of the Department of Education, (Gabriel, Anderson, Benson, Gordon, Hill, Pfannensteil, and Stonehill, 1985), also assessed the sustained achievement of program participants. Their study relied on standardized

---

[25]These summer patterns indicate only relative loss, not absolute loss. The scale re-calibrates scores on each testing occasion, so that the average score is always 100. Thus, a horizontal line indicates that students are gaining at an average rate, and a downward line indicates that students are gaining less.

achievement test data collected in spring 1982 and spring 1983 in 17 school districts or State educational agencies. Analysts reviewed achievement patterns based on students' participation in Chapter 1 during these two school years, with some students participating in Chapter 1 during both years, some during neither year, and some during the first year only or the second year only. Although not nationally representative, the study included over 66,500 second- through sixth-grade children.[26]

Student achievement patterns resulting from this analysis are presented in Figure 4.7. The test scores are re-calibrated for each test administration, so that normal growth yields a horizontal line. The four lines displayed in Figure 4.7 indicate student performance during the second year only. That is, the first data point shown marks student achievement at the end of the first year. The patterns in Figure 4.7 are similar to those reported by the Sustaining Effects Study: students who were never in Chapter 1 had higher, relatively stable achievement scores over time. Those who participated during both years had the lowest scores, though they showed small gains during this period of participation in Chapter 1. Those who participated during one of the two years scored in between these other two groups. If they participated during the first year and not during the second, they exhibited declines during the second year; if they participated during the second year and not the first, they exhibited gains during the second year.

The researchers responsible for this study hypothesized that there are three distinct populations of students whom school districts serve: (1) the general population, (2) a population needing some remedial assistance, and (3) a population needing considerable compensatory assistance. The uppermost line in Figure 4.7

---

[26]The study sample had slightly higher pretest scores (ranging from 1.4 to 4.3 NCEs in reading and from 0 to 2.2 in mathematics) than those reported in TIERS for the overall Chapter 1 population. This may be due to special features of the California Achievement Test which is used, or to differences in the populations served by districts participating in this study.

Figure 4.7

Achievement of Students by Pattern of Participation in Chapter I Across Two Years, 1982-83

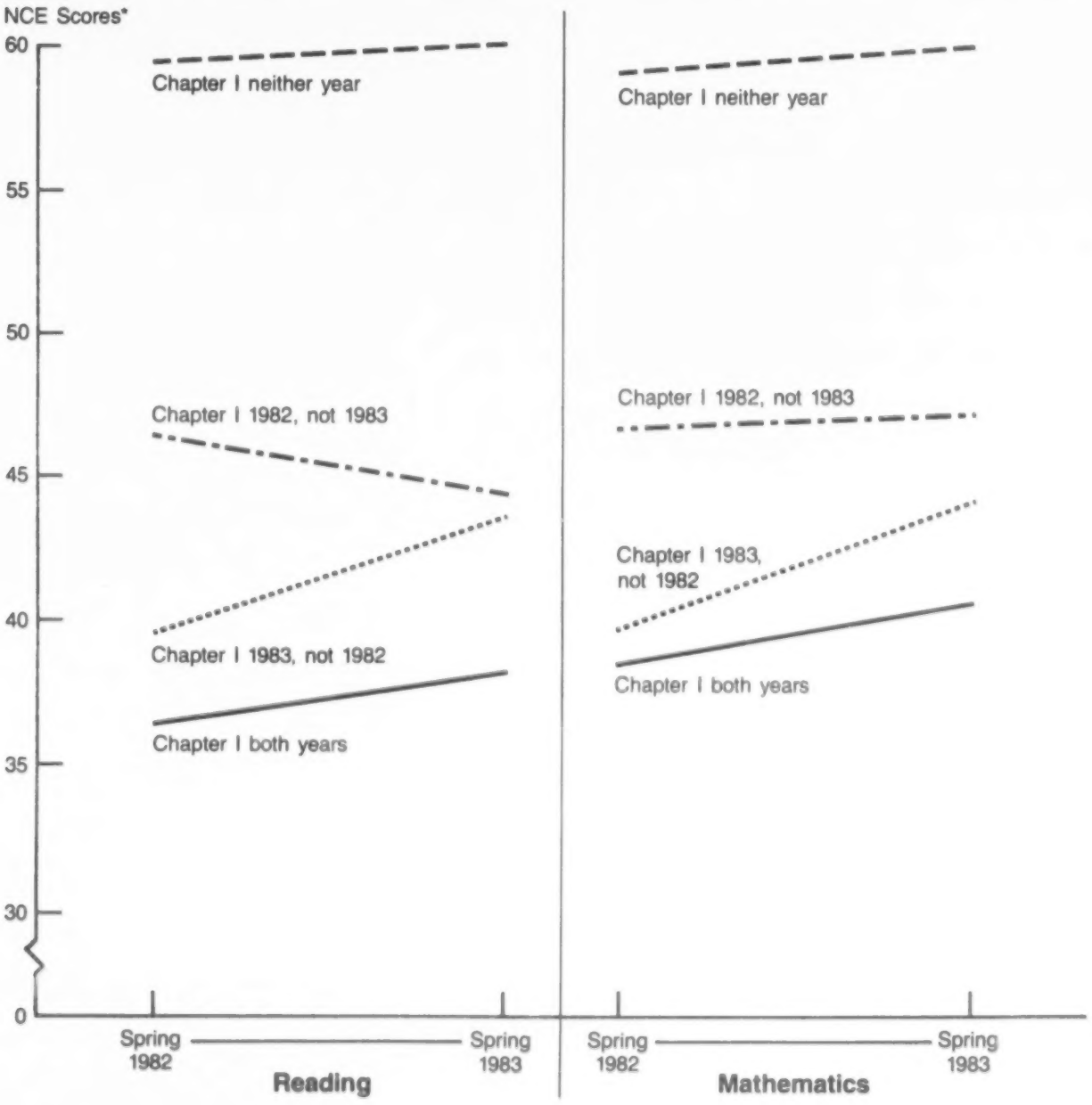


Figure reads: From the spring of 1982 to the spring of 1983, students who received no Chapter I services in either school year had average NCE scores of slightly less than 60, and demonstrated slight gains between spring 1982 and spring 1983 in both reading and mathematics.

*NCE scores are transformations of percentile ranks to a standardized equal-interval scale; see Appendix E.

Source: R. Gabriel, B. Anderson, G. Benson, S. Gordon, R. Hill, J. Pfannenstiel and R. Stonehill, *The Sustained Achievement of Chapter I Students*. U.S. Dept. of Education, January 1985.

represents the achievement growth of the first group; the middle two lines represent the second group, and the lowest line represents the third group. The group that received continuous compensatory education appears to reap the smallest achievement benefit; yet we cannot know what their achievement patterns would have been in the absence of Chapter 1 assistance.

Finally, recent longitudinal analyses of the achievement levels of former Chapter 1 participants in Columbus, Ohio (Amorose, Brown, Duffy, Morgan, and Thompson, 1986) and in St. Louis, Missouri, and Lincoln, Nebraska (Pfannensteil, 1986) compared the fourth- and fifth-grade achievement scores of students who had received Chapter 1 services for one, two, three, or four of the preceding years. These data indicated that many groups of students were able to retain standardized achievement scores at approximately the levels they had at the close of their Chapter 1 services. However, students who received three and four years of service tended to have relatively lower scores even when they left Chapter 1 and had still lower scores when retested later on.

Long-Term Program Effects in Areas Other Than Achievement Test Scores

Recently, several analysts and policy makers have taken an interest in other indicators of program benefits, such as long-term patterns of school attendance, grade retention, graduation rates, and even future education and career accomplishments. Assessing long-term program effects would require researchers to follow a group of students for several years. Even if a study were able to follow a sample of students, it would encounter interpretive problems in attempting to account for differences. The longer the period of time between the receipt of services and the measurement of outcomes, the less able analysts are to attribute observed changes to participation in a program: too many things have happened in the meantime.

While such long-term studies have not been conducted in the context of Chapter 1, they have been conducted for preschool programs. The evidence suggests that, while achievement-score gains tend to fade away over time, other important outcomes, such as school attendance rates and grade retention rates tend to remain (Smith, 1986). The discrepancy between these long-term behavioral patterns and the long-term achievement pattern make the data difficult to interpret.

In an effort to disclose the long-term education experiences of former Chapter 1 students, the National Assessment of Chapter 1 funded several small projects to analyze existing State and local data bases. Most were unable to follow up on students for more than five years, and most concentrated on long-term achievement patterns. These studies indicated that lower-scoring pupils were more likely to receive multiple years of Chapter 1 services (Amorose, et al., 1986; Pfannensteil, 1987, Kirshstein, 1986; Plato, et al., 1986) and that they were more likely to receive services from other categorical programs as well (Plato, et al., 1986; Pfannensteil, 1986). Where data on grade retentions were available, they indicated that Chapter 1 participants were retained in grade more often than non-participants (Amorose, et al., 1986). Former Chapter 1 students in Montgomery County, Maryland were less likely than other students in that district to pass the State's ninth-grade functional reading and mathematics tests, and more likely to have been suspended in the past three years. These data suggest that, even after leaving Chapter 1, these students are likely to continue to have problems in school.

Summary and Discussion

In this chapter we reviewed evidence regarding the growth of disadvantaged students' achievement over the summer and into subsequent school years. The evidence indicates that the achievement test scores of disadvantaged students tend to decline, relative to those of more advantaged students, as they progress through the grades,

and that the achievement gap widens during each summer. Chapter 1 assistance during the school year raises the relative achievement levels of some of its students and helps others maintain their relative position. Once students leave Chapter 1, their scores again decline. Because participation in Chapter 1 is related to achievement levels at the outset, it is difficult to estimate what would happen if students continued to receive services over many years. Our principal findings are as follows:

7. The achievement gap between disadvantaged and advantaged students appears to widen during the summer months.

Though few studies have been designed specifically to monitor the gap between the achievement levels of advantaged and disadvantaged students, available data sources indicate both that the gap widens as students move through the grades and that the summer months contribute to the increase. By the time students are in sixth grade, low-achievers are substantially behind their higher-achieving counterparts. Further, the amount students gain during the school year is much less, relative to the differences between high- and low-achieving students, than gains made by younger students.

8. Title I-supported summer programs have not narrowed the achievement gap; however, these programs often were not designed to be academically rigorous.

In earlier periods of Title I history, summer school programs were a popular option, in part because it was believed that these might prevent the achievement of disadvantaged students from falling even farther behind. Though the evidence suggests that such programs did not substantially alter student achievement, it also indicates that these programs were not designed to be academically rigorous. Further, they usually lasted only six weeks. Thus, the extent to which such programs may benefit disadvantaged students is still not known.

9. Students who discontinue Title I appear gradually to lose ground when they no longer receive services.

When Title I was first legislated in 1965, many observers had assumed that a year or two of compensatory education services would be sufficient to alter the educational futures of disadvantaged children. The data suggest that such long-term benefits do not occur. There are at least three hypotheses that can account for these findings.

One hypothesis is that gains appear not to be sustained because the content children learn each year--and the content that achievement tests measure each year--differs from the preceding year's instructional content. Chapter 1 may help children master their second-grade material, for instance, and children may sustain that mastery through future grades. But second-grade assistance would not help children master their fourth-grade material. Under this line of reasoning, children are in fact sustaining the gains they made while participating, but the achievement tests of future years no longer measure that content.

A second, and related hypothesis is that disadvantaged children continually learn at a slower rate, and that they consequently continue to need more instruction and repetition in order to learn and to retain what they have learned. This hypothesis implies that most Chapter 1 students would need to receive services every year, from kindergarten through twelfth grade, not just for one or a few years.

A third hypothesis is that Chapter 1 students need help learning to learn. This hypothesis suggests that it may indeed be possible to provide a short-term supplementary service which will have a long-term impact. But the nature of that service would be quite different from what is now typically provided under Chapter 1. In fact, under this hypothesis, Chapter 1 instruction which focuses on curriculum content will necessarily need to be repeated year after year, for it does not teach students how to learn from the regular program.

10. Chapter 1 students with very low achievement levels appear to maintain their relative achievement levels while participating in Chapter 1, but not to move ahead. However, evidence suggests they would have lost ground relative to their peers if they had not received services.

The evidence available suggests that the lowest-achieving students receive multiple years of service, and that, while their achievement scores do rise from year to year, they do not rise enough to substantially alter the students' relative academic standing. Thus they appear to continue on at relatively low achievement levels.

However, because the learning curves of low-achieving students differ from those of higher-achieving students, it is difficult to estimate the extent to which, if at all, Chapter 1 services have benefitted their lowest-achieving participants. Evidence presented in the last chapter suggests that these students could have fallen farther behind their peers if they had not received Chapter 1 services. If this is so, then what appears to be a lack of impact, because students retain their relative position, may in fact represent an impact in that students have not fallen farther behind.

11. No nationally-representative studies have examined the long-term effect of Chapter 1 programs on graduation rates, future education, or adult literacy.

An assumption underlying the Title I and Chapter 1 legislation is that compensatory education will help poor children improve their future educational prospects. While research indicates that some Chapter 1 students continue to experience a range of difficulties as they proceed through school, information about the long-term effects of participating in Chapter 1 programs is not available. Studies of these effects would be difficult to conduct and to interpret if they were conducted, both because participation in Chapter 1 can vary greatly from student to student, and because it would be difficult to keep track of former students and their educational experiences over a long period of time.

5. Project Characteristics and Student Achievement

Central Findings

12. **Large-scale studies designed to identify particular project characteristics that improve student achievement test scores have yielded inconsistent or inconclusive findings.**

13. **Researchers have recently identified a number of instructional practices that are likely to increase the achievement of disadvantaged students and that can be used in Chapter 1 programs**

Chapter 1- and Title I-supported compensatory education services have been provided to disadvantaged students for over 20 years. Throughout that time period, questions have been continually raised about the value of these programs. Program critics point to the over $3 billion spent each year for these services, and argue that the gains we described in earlier chapters are not sufficient to justify that cost. Program advocates point to the serious need that disadvantaged children have for educational help. Most of the studies we reviewed in earlier chapters were motivated by these debates and have contributed to them. Evidence on long-term program effects and on the learning rates of different kinds of children suggests that the problems of educational disadvantage are much more difficult to solve than the original designers of Title I legislation had assumed. Evidence of one-year program effects, on the other hand, suggest that these services do noticeably alter students' achievement in mathematics and reading.

Even more problematic, at least in regard to this debate, is the apparent variability of program effects across subject matters, grade levels, types of student served, and localities. Such variability suggests that it is at least hypothetically possible for Federally-supported services to have a greater average impact than they now have. If services could be improved at the lower end of the distribution of

student achievement, so that they emulated the upper end, perhaps the net impact of the legislation would be increased.

In an effort to improve the effectiveness of Chapter 1 services, researchers and practitioners have launched many efforts over the last 20 years to identify effective compensatory education practices. They have sought such practices primarily in order to assist other, less effective projects. Efforts to spotlight effective practices have taken three forms: (1) identifying effective Chapter 1 projects and informing other districts about them, (2) identifying specific project features that seem to promote student achievement and that could be implemented in other projects; and (3) identifying effective practices in general education that could be implemented within Chapter 1. This chapter reviews each of these efforts.

Efforts to Identify Effective Projects

The earliest Federal attempts to improve Title I programs concentrated on identifying and describing local projects that demonstrated exceptional achievement increases, in the hope that other school districts could adopt these projects and thereby increase the effectiveness of their own programs. These efforts prompted development of a dissemination network and publication of an annual catalogue of effective practices, called Education Programs that Work. The annual catalogues described effective projects supported by other Federal categorical programs as well as Title I.

This method of dissemination proved popular, in part because of the recognition that local project developers received. However, its success as a mechanism for improving Chapter 1 projects has been mixed, for two reasons. First, it was not possible to know which aspects of effective projects are responsible for students' achievement gains. A project could be successful because of the way it involves parents, the way it uses teachers or teacher aides, the way Chapter 1 services are

coordinated with the regular school program, or for other reasons. Indeed, its success may be the result of one talented teacher. Second, as dissemination of effective projects became more commonplace, it became clear that districts did not adopt them in their entirety. Instead, districts adapted them to fit local circumstances. The combined effect of, on the one hand, researchers disseminating projects whose most important features could not be identified, and of, on the other hand, districts adapting these projects to fit their own circumstances, raised questions about the ultimate effectiveness of this improvement strategy. For districts could unwittingly alter the very feature that contributed most to student achievement, and no one would know it. Thus, though the recognition of effective projects provides an incentive for projects to do well, it is less useful as a strategy for improving the effectiveness of other projects, in part because such recognition of effective practice does not improve our knowledge of what contributes to effectiveness.

Efforts to Identify Effective Features of Chapter 1 Projects

One strategy for increasing our knowledge of what contributes to effectiveness is to study a wide range of projects, not just effective projects but ineffective projects as well, and to see how these projects differ. Three national data sources have been used to study the relationship between different project features and gains in student achievement: the Instructional Dimensions Study, the Sustaining Effects Study, and TIERS.

The Instructional Dimensions Study

To sort out the relative benefits of particular project features, NIE's Instructional Dimensions Study (NIE, 1976; Cooley and Leinhardt, 1980) developed a model of the instructional process which focused on five factors: student motivation, opportunity to learn in the classroom, degree of individualization in the program, the nature of instructional events, and the characteristics of project teachers. These factors were

further broken down into a number of specific descriptors. Using these descriptors as the basis for their data collection, researchers observed 400 first and third grade classrooms in 14 school districts, testing students in each classroom. Their goal was to estimate the contribution of each instructional factor to student achievement.

The Instructional Dimensions Study yielded three important findings. First, it found that individualized instruction did not make a difference in student achievement. That is, students gained approximately the same amount whether they were taught individually or in groups. This finding was particularly important when it was published in 1976 because Congress was considering whether to require individualization under Title 1. The Instructional Dimensions Study defined individualization to mean that: (1) specific learning objectives were assigned to individual children, (2) children were taught in small groups or were individually paced, (3) children had the benefit of individual diagnosis and prescription, or (4) there were alternate learning opportunities and sequences available for individual children.

Second, it found that pullout programs were a more effective instructional arrangement in some circumstances, and in-class programs were more effective in others. First-grade students seemed to benefit more from in-class programs, whether the subject was reading or mathematics. Third-grade students, on the other hand, derived more benefit from pullout programs in mathematics and showed no differences in reading. The extent to which pullout and in-class configurations may have yielded other patterns of effects in other grades could not be determined. Because of ongoing debates about program structures, this finding continues to be important today.

The third important finding from the Instructional Dimensions Study was that the factor called opportunity to learn was the only one consistently and significantly associated with student achievement. This factor included such program features as the amount of time devoted to instruction and the size of the groups in which children

were taught. However, it also included the extent to which the content of the achievement test matched the content of the Chapter 1 curriculum, and further analysis indicated that this content overlap between test and text was the most important aspect of students' "opportunity to learn."[27]

The Sustaining Effects Study

The Sustaining Effects Study also examined the relationship between classroom instructional practices and student achievement. Unlike the Instructional Dimensions Study, the Sustaining Effects Study gathered data on instructional techniques from teacher questionnaires rather than direct observation. Analysts in this study assessed the relative contributions of many academic and nonacademic factors to students' achievement and tested a number of models of the educational process. In each case, the outcome of interest was student achievement scores (Carter, 1983).

The Sustaining Effects Study researchers also found evidence of a relationship between instructional time and achievement (Wang, 1981), though the relationship did not appear in every grade and subject examined. Furthermore, time spent specifically with a special program teacher or in a special setting did not have appreciable effects on achievement growth, though time spent in regular instruction did. Other factors that appeared to make a difference were the teachers' years of experience, the frequency with which students received feedback on their performance, and the lack of disruption in the classroom. Sustaining Effects Study researchers concluded that the amount of compensatory instruction would not by itself close the achievement gap between compensatory education students and their nondisadvantaged peers and that

---

[27]This does not necessarily mean that teachers were or should be "teaching to the test." Content overlap means that, for instance, if teachers are teaching Roman numerals in their mathematics class, the test includes questions on Roman numerals. Or, if the reading textbook includes a large nature vocabulary, so does the reading test. Freeman et al (1983) have shown that content overlap varies substantially across different combinations of textbooks and standardized tests.

more information was needed about the components of compensatory instruction to know which are most important to achievement.

TIERS

TIERS included a reporting form which States could use to describe specific project characteristics, as well as student achievement gains, in grades 2, 6, and 10. Project characteristics included the number of days the typical student participated in the Title I program, whether the program was offered in class or on a pullout basis, and the student-to-teacher ratio.

For the most part, State educational agencies' reporting of programmatic data has been sporadic and prone to error, more so than has been true of the TIERS achievement data. Analyses of the data have yielded no clear patterns relating project characteristics to gains in student achievement (Wood, 1984).

Analysis of Efforts to Identify Effective Project Features

Despite elaborate and sophisticated efforts to identify effective project features, only a few such features have been found. Four important limitations account for these results.

The first concerns the ability of researchers to know in advance which features of an educational program are likely to be significant and should therefore be documented in their study. The Instructional Dimension Study researchers relied on a theoretical model of the instructional process to decide what to observe, and then observed actual practice to document the factors they expected to be relevant. The Sustaining Effects Study researchers relied on existing student records and on teacher questionnaires for their evidence of practice. In both the Sustaining Effects Study and TIERS, the characteristics that tended to be documented were those that could be efficiently observed and measured (e.g., whether the teacher has an aide, the teacher's years of experience, whether instruction occurs within the regular classroom or in

another setting). Yet what may be most significant is not the presence of a classroom aide but the training the aide has received or the amount of time the teacher spends discussing each day's instructional goals with the aide. Similarly, the physical location of services -- pullout versus in class, for instance -- may not be as important as the character of the services provided. When the Instructional Dimensions Study was completed, the NIE convened a group of teachers who had participated in the study to discuss its findings. In that conference (NIE, 1978), teachers described a wide variety of practices that were all labeled "pullout" but that had remarkably different instructional meanings. They argued that the distinction between pullout and in class configurations was not as instructionally important as these other variations were.

Second, this line of research is complicated by the fact that Chapter 1 services, while defined as if they were entirely separate from the rest of the school program, are not separable in their impact. In most cases, Chapter 1 services are provided during only a small portion of the school day--for half an hour, perhaps. During the rest of the school day, Chapter 1 students receive the same instruction as other students do. And their achievement is influenced by both sets of instruction. The Instructional Dimensions Study concentrated only on the Chapter 1 portion of the school day, and attempted to determine aspects of Chapter 1 services that were related to gains in student achievement. Yet at least some of the students' achievement growth must have been due to the regular program. To the extent that student achievement is also influenced by the regular program, the researcher has more difficulty separating out important features of Chapter 1 projects. The Sustaining Effects Study researchers included measures of both Chapter 1 and regular instruction in their analysis. When students received instruction from multiple teachers, the "practice" was defined as the average of these teachers' responses to the questionnaire. Consequently, when these researchers identified a factor such as "lack of disruptions in

the class" as relevant to student achievement, they were referring to a factor that matters regardless of whether it occurs within Chapter 1 or in the regular setting. Consequently, their strategy also makes it difficult to identify the important features of Chapter 1.

Third, these studies have tended to search for the unique effects of each project's characteristics, even though in practice these characteristics operate in tandem to influence student achievement. For example, classroom aides may contribute significantly to student achievement in one project and yet contribute little in another project. The difference in the salience of their roles may be related to other project features such as the curriculum or student grouping practices. Yet because the contribution of classroom aides varies from one project to another, their average impact can appear to be nil. The number of ways in which these instructionally-relevant project characteristics may be combined exceeds the capacity of existing data to identify, define, and assess these variations. As we have seen, however, attempts to identify whole projects that are effective have been equally unsatisfying.

Finally, these analyses are complicated by the variety of relationships that can exist between project features and student needs. "Chapter 1 children" do not constitute a homogeneous group, and the features of local Chapter 1 services could purposely vary in response to student needs. In our first report, we showed that the population considered to be educationally deprived can vary a great deal from school to school because the overall populations served by these schools differ. Further, even within a given school or district, Chapter 1 children are not a permanent subset of the student population. Instead, they come and go as their achievement fluctuates relative to that of their peers. These variations in populations served, combined with variations in the characteristics of Chapter 1 services across schools and districts,

make it extremely difficult to determine whether a particular kind of services is generally efficacious.

Efforts to Identify Generally Effective Practices Not Limited to Chapter 1

Large-scale studies designed to identify effective features of Chapter 1 projects have increased our knowledge more than have project recognition efforts. But these efforts are large and expensive undertakings, and the knowledge gained has not been sufficiently rich or informative to guide local practice. Two recent efforts to improve our understanding of the relationship between practices and student outcomes draw more on existing research from education in general, rather than confining themselves to the Chapter 1 context. These efforts have sought to identify educational practices generally recognized as effective, whether or not such practices are used in Chapter 1. One of these was conducted by the Education Department as a program improvement effort, the other by the National Assessment of Chapter 1, specifically for this report.

ED's Synthesis of Factors Promoting Chapter 1 Improvement

ED's most recent program improvement effort in Chapter 1 incorporates both the project recognition approach and the effective-characteristics approach. It began with a review of research on the characteristics of effective schools. In the past 10 years or so, a substantial body of research has accumulated regarding the features of effective schools, and several researchers have attempted to synthesize these findings and to develop a list of school characteristics that appear to make a difference. ED capitalized on these findings and on researchers' interests in codifying the findings. It created a list of characteristics that were particularly relevant to Chapter 1 projects. ED's list included such organizational attributes as a "positive school and classroom climate" and "parental/community involvement"; and it included such instructional attributes as the "use of academic learning time". Using these characteristics as a

guide, ED then asked State educational agencies to nominate Chapter 1 projects that exemplified these features, and asked the nominees to describe how their programs incorporated these ideas. From these documents, ED identified 113 Chapter 1 projects as being particularly effective. Finally, ED described, on one hand, the school characteristics linked in the research literature to higher average achievement scores and, on the other, its examples of Chapter 1 projects which implemented these ideas. ED produced a two-volume Sourcebook (Griswold et al., 1986) which has been disseminated to Chapter 1 projects across the nation as a guide to improving project operations.

This approach combines the best features of the project recognition approach and the effective-features approach. With these descriptions of specific projects, other projects are given concrete examples that they can emulate. In addition, the effectiveness of the projects is analyzed, for they are defined according to the features that ED expects, based on research to be important. Finally, the research findings regarding these important features is also summarized, so that adopting projects may learn the principles of effective practices that are exemplified by these projects. Thus, if they choose to adapt the projects to suit their local circumstances, they may do so while still honoring the important principles underlying the project.

One difficulty with this approach, however, is that the effective practices were derived from research on schools, not on Chapter 1 projects. Some of the findings from this body of research may be particularly relevant to Chapter 1 projects while other findings may not be. Academic learning time, for instance, is something that many researchers have found to make a difference in a variety of settings, and is likely to matter in the context of Chapter 1 as well. But other school features identified by that research, such as a positive school climate and a set of goals that are shared by faculty and administrators within the school may not transfer to

Chapter 1. School climate and shared instructional goals both refer to the culture of the school as a whole, not to particular programs within the school. If Chapter 1 programs achieved these features while the rest of the school did not, the presumed benefit to student achievement may not occur. In fact, some researchers (e.g., Smith, 1986) have argued that because Chapter 1 programs involve only a subset of teachers and a subset of students within each school, they may actually encourage separate subcultures with separate educational goals, thus decreasing the school's ability to develop the sort of climate now recognized as important to student achievement.

The National Assessment's Review of Effective Educational Practices

The National Assessment of Chapter 1 asked a number of researchers to review the research evidence with which they were most familiar and to describe the characteristics of sound educational practices that could or would be appropriate for disadvantaged children.[28] Researchers were asked to review evidence about effects of practices in five areas (Williams et al., 1987):

- Chapter 1 school and student selection procedures;
- Chapter 1 program design and staffing structures;
- Relationships between Chapter 1 and regular programs;
- Parent involvement; and
- Curriculum and instruction practices.

These five topics consider Chapter 1 both in isolation and in conjunction with the regular school program. The first two topics address specific features of local Chapter 1 projects, and most of the research reviewed on these topics came from evaluative studies of Title I and Chapter 1. The second two topics--relationships between Chapter 1 and the regular program and parental involvement--have more to do

---

[28]These researchers were not asked to address the political or fiscal implications of their findings but instead to state their interpretations of current research findings.

with the child's entire educational experience, and where and how Chapter 1 services fit into the overall configuration. Finally, the last topic--curriculum and instructional practices--addresses issues that are almost completely independent of Chapter 1, though the findings in this area could be implemented within Chapter 1 as well as in any other setting.

The researchers who addressed the issue of Chapter 1 school and student selection discussed our first interim report to Congress at length, and suggested that the provision of services only to the lowest-achieving students within Chapter 1 schools may not be the best way to allocate service. Rather, these researchers suggested targeting Chapter 1 funds on those schools serving the highest concentrations of low-income students, and then serving all students within these schools.

There were two reasons for this suggestion. The first reason was that Chapter 1 now serves many students attending schools which do not serve unusually high concentrations of children from low-income families. These researchers thought that concentrating funds in fewer schools would result in the provision of services to a more needy group of students. In our first report to Congress, we showed that students attending schools with high concentration of poor children were doubly disadvantaged: even non-poor students attending such schools are more likely to be low achievers than are non-poor students in other schools. The second reason for this suggestion, however, had to do with the research relating to effective schools. One researcher (Smith, 1986) argued that concentrating services in fewer schools offered the additional benefit of enabling schools to develop schoolwide projects, thus eliminating the potentially divisive subcultures now hypothesized to exist in Chapter 1 schools; and that concentrated services and schoolwide projects would also eliminate

cumbersome within-school accountability mechanisms that now accompany Chapter 1 funds.

The evidence suggests that concentration of funds would indeed increase the number of poor and low-achieving students served. However, the relationship between school and student selection decisions on one hand, and program quality on the other, is not clear-cut. If the Chapter 1 funds in a schoolwide project were used to add a science teacher to the faculty, they may not increase the students' academic learning time. On the other hand, if funds were used to train all teachers in the school about the special educational needs of disadvantaged students, to facilitate an organizational climate and a set of high expectations for student learning, then perhaps such an alteration would positively benefit students.

With regard to program and staffing structures, researchers expressed frustration at the lack of knowledge about such basic practices as serving students within their own classes versus pulling them out of these classes to receive services in another setting. As we pointed out earlier, the educational setting may be less important to student learning than is the instruction that occurs within the setting. One researcher concluded: "What is disconcerting. . . is that the proponents of pullout have comparatively little evidence to justify its widespread use and the critics of pullout would be similarly hard-pressed to justify a wholesale shift to in-class instruction" (Archambault, 1986 p. 2).

Another program and staffing feature that has been debated considerably in the past is the ratio of students to teachers. Most of education occurs in traditional classrooms, with class sizes ranging from 20 to 35 students, depending on grade level. The available evidence suggests that differences in this range do not substantially influence student achievement. However, the researcher at this conference who addressed this issue felt that group sizes as low as those adopted by Chapter 1

programs--often as few as 10 students in a class--do make a difference (Cooper, 1986). Thus, one aspect of Chapter 1 that could be contributing to student achievement is the reduced student-teacher ratios which it supports. Such small groups may enable Chapter 1 teachers to provide students with the individualized help and attention they need.

Yet these ratios are only provided for a short period each day. For the bulk of the school day Chapter 1 students are in regular classrooms and share their regular teachers with 30 other students. This fact raises questions about the role Chapter 1 services play in the entire school program, a topic also addressed by conference participants.

Two important points were made about the relationship between Chapter 1 and the regular program. First, Allington and Johnston (1986) argued the importance of "curricular congruence" for improving student learning. Second, Griffin (1986) argued the importance of coordinating inservice training for teachers. Both pointed to evidence of shared instructional goals as an important feature of effective schools. Griffin pointed to evidence that staff training is more effective when it is provided to all teachers in a school, rather than to subgroups such as Chapter 1 teachers, and Allington and Johnston pointed to evidence that learning is impeded when students receive inconsistent lessons. When teachers do not coordinate, the responsibility for finding connections across lessons falls on the students, and students who are having difficulty in school to begin with have a particularly difficult time when they must also take responsibility for coordinating their own instruction.

With regard to parent involvement, there was consensus among these researchers that student achievement improves when parents take a strong interest in their children's education and encourage them to do well (McLaughlin and Shields, 1986; de Kanter, Ginsburg and Milne, 1986; Rich, 1986). Such attitudes may manifest themselves

through such behaviors as helping children with their homework or attending parent-teacher conferences. Some forms of assistance need not rely on high levels of parent education: turning off the television when it is time to do homework and providing a place for students to work can convey the importance of these activities to students.

Finally, with regard to curriculum and instructional practices, researchers identified a number of instructional factors which, taken together, increase students' opportunities to learn.[29] One researcher summarized the evidence as follows: achievement scores are improved when teachers "emphasize academic instruction, expect their students to master the curriculum, and allocate most of the available time to curriculum related activities" (Brophy, 1986 p. 8). Other researchers not participating in this conference have also pointed out that time devoted to instruction makes a difference to students' achievement, but that time alone is not sufficient to raise achievement. The time must be used well (Karweit, 1983; Walberg and Frederick, 1985; Cooper, 1986). Our researchers also argued that when students are actually engaged in academic work, rather than getting organized or moving from one activity to another, and when teachers teach actively, rather than assigning large amounts of seat work, students learn more.

Researchers disagreed, however, on the extent to which teachers should alter their behaviors when working with disadvantaged students. One review indicated that disadvantaged students "need more control and structure from their teachers: more active instruction and feedback, more redundancy, and smaller steps with higher success rates. This will mean more review, drill, practice, and thus more lower-level questions" (Brophy, 1986 p. 35). Brophy also acknowledged that "across the school

---

[29]One reason this factor may have been significant in these studies, but not in the earlier large-scale studies described above, is that these researchers were able to code subtler and more instructionally-relevant aspects of opportunity to learn. These researchers agreed, for instance, that the time allocated to instruction would not, by itself, increase the time students spent actually engaged in learning.

year, it will mean exposure to less material, but with emphasis on mastery of the material that is taught and on moving the students through the curriculum as quickly as they are able to progress" (Brophy, 1986 p. 35).

Other researchers (Romberg, 1986; Calfee, 1986) argued, however, that such special treatment of disadvantaged students could translate into lower teacher expectations, less material covered, a more boring educational experience for students, and, ultimately, lower achievement. These authors argued that an exclusive focus on basic skills, especially in later grades, would not allow students to go beyond these skills to grasp important concepts and to think and solve problems.

Summary and Implications for Chapter 1

A number of important findings emerged from this collection of research reviews. Some of these findings suggest hypotheses for why Chapter 1 services have the outcomes they do. The research on student-teacher ratios, for instance, suggests that Chapter 1 may facilitate learning by providing small class sizes, even though these smaller groupings occur for only a portion of the school day. On the other hand, research on the features of effective schools suggests that Chapter 1 may also hinder student achievement by restricting the school's ability to create the shared academic goals, high expectations, and strong achievement-oriented school culture that are now recognized to be important to student achievement. These findings also suggest that disadvantaged students might learn even more if, for instance, the sizes of their regular classes were reduced substantially, or if Chapter 1 teachers were more fully incorporated into the school's overall instructional program.

Other findings indicate that the features of Chapter 1 programs that have tended to capture the attention of policy makers--features such as pullout and in-class programs and individualized instruction--are not particularly influential on student achievement, and should probably not receive the amount of policy concern that they

traditionally have received. Still other findings point to areas outside the influence of Federal policy makers--the curriculum, the teachers' instructional strategies, the effective use of learning time in classrooms, and the culture of the school as a whole.

One feature of Chapter 1 for which these findings are relevant is the supplementary character of Chapter 1 services. The research reviewed here indicates that the more opportunities students have to learn, the more they actually learn, and Chapter 1 programs are generally designed to provide students with additional instruction in reading and mathematics. But as a supplement to the regular program, Chapter 1 services usually represent only a portion of a child's total instructional day. The most recent evidence suggests that Chapter 1 services average 44 minutes a day for four days a week (Advanced Technology, 1983). Further, Chapter 1 services may not actually provide additional opportunities to learn, for two reasons. First, students usually receive Chapter 1 services during the same six-hour day and nine-month year that are used to serve non-participating students. To the extent that time allocated to Chapter 1 is taken away from other instruction, the presumed additional instruction provided by Chapter 1 may not be an addition at all but merely a replacement of one form of instruction for another,[30] though presumably these new services are designed to be more intensive than the regular services. Second, opportunities to learn can be greatly increased by increasing the efficiency of regular classroom instruction and by keeping students engaged in learning. Because Chapter 1 services are supplementary, Chapter 1 can do little to improve the regular instructional program, and may in fact introduce inefficiencies as students are grouped and re-grouped, or as they move back and forth from their regular program to their Chapter 1 program.

---

[30] This need not be a violation of the supplement-not-supplant rule, for that rule refers to fiscal rather than educational relationships among programs.

Even if Chapter 1 provided superior teachers to those that students otherwise receive--and there is no reason to believe that this would be so--the quality of Chapter 1 services may not be sufficiently better to compensate for poor teaching during the remainder of the day. Regardless of the quality of services provided during the Chapter 1 portion of the day, students still spend five and a half hours a day in the regular program, and the quality of that program will influence their ability to learn.

The supplemental character of Chapter 1 is also implicated by the research evidence regarding the importance of schoolwide goals and program coherence. To the extent that the supplemental nature of Chapter 1 creates a separate cadre of teachers with duties, goals and responsibilities different from those the regular teachers have toward Chapter 1 students, Chapter 1 may diminish the coherence of the total educational program that students receive. The research findings suggest that Chapter 1 teachers and services should not be isolated from the rest of the school program--some have even suggested that the legislation should not restrict the use of Chapter 1 funds to supplemental services. Without such a restriction, the argument goes, funds could be used to improve the regular program. But it is also true that, without such a restriction, students may lose benefits, such as smaller student-teacher ratios, they now receive from Chapter 1.

Another feature of Chapter 1 for which these findings are relevant is its provision for parent involvement. Chapter 1's predecessor, Title I of the ESEA, required parent advisory councils and provided considerable detail regarding the nature and role of these councils. These provisions were curtailed in Chapter 1. While parent involvement in the governance of local projects may serve other important goals, the research reviewed here suggests that parents' direct involvement in their own children's education is more likely to have a positive effect on student

achievement. Such involvement occurs when parents encourage their children to do their homework, when they attend school and observe classes, and when they tutor their children, for instance.

But such involvement may be difficult to accomplish through Chapter 1 for two reasons. First, Chapter 1 constitutes only a portion of the child's school day, and it is not considered to be the main educational program. Traditional mechanisms for obtaining parent involvement, such as report cards or parent-teacher conferences may not be sufficiently flexible to incorporate Chapter 1 instruction. Second, to the extent that Chapter 1 parents are themselves less educated, they may feel intimidated by educational institutions and be reluctant to participate in their children's education. Further, the quality of parent-school relations may vary considerably from district to district, so that no general rule of thumb would necessarily apply. Therefore, the implications of these findings for determining whether or how the Chapter 1 legislation could facilitate this form of parent involvement is not clear.

Developing an optimal role for Chapter 1 to play in the variety of schools in which it now exists is extremely difficult. Yet it is clear that this role, and its relationship to the regular program, makes a difference to students. Schools may try to coordinate Chapter 1 and the regular program by using common tests or common textbooks, by placing Chapter 1 teachers or aides under the supervision of regular teachers, by maintaining joint record-keeping systems on student progress, and so forth. The variety of practices is probably so great that Chapter 1 services sometimes enhance the rest of the school program, sometimes hinder it, and sometimes exist beside it but completely independent of it.

Without more knowledge about what actually happens in Chapter 1 classrooms and in regular classrooms, we cannot know which of these research findings--or combinations of findings--best explain the achievement data we have reviewed here, or

which suggest the most fruitful directions for increasing the effectiveness of Chapter 1 programs. The National Assessment of Chapter 1 is currently completing several studies designed to describe the services currently provided under Chapter 1, including examinations of how these services relate to the regular program, whether they are consistent with research on educational practices for disadvantaged students, and the extent to which they are influenced by Federal as opposed to State or local policies. These findings will be presented to Congress in our final report.

References

Advanced Technology, Inc.
1983 Local Operation of Title I ESEA 1976-1982: A Resource Book. McLean, VA: Author.

Allington, R. L. and P. Johnston
1986 "The Coordination Among Regular Classroom Reading Programs and Targeted Support Programs." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Amorose, R. A., R. D. Brown, J. A. Duffy, K. L. Morgan, and G. Thompson
1986 Analysis of School District Records to Study the Effectiveness of Chapter 1 programs Over a Five Year Period. Columbus, OH: Columbus Public Schools.

Archambault, F.
1986 "Instructional setting: Key Issue or Bogus Concern." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Brophy, J.
1986 "Research Linking Teacher Behavior to Student Achievement: Potential Implications for Instruction of Chapter 1 Students." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Calfee, R.
1986 "Curriculum and Instruction: Reading." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Carpenter, M. A. and P. A. Hopper
1985 Synthesis of Chapter 1 Data: Summary Report (Prepared for State and Local Grants Division, Office of Planning, Budget and Evaluation, U.S. Department of Education). Reston, VA: Advanced Technology, Inc.

Carter, L. F.
1984 "The Sustaining Effects Study of Compensatory and Elementary Education." Educational Researcher 13(7).

Carter, L. F.
1983 A Study of Compensatory and Elementary Education: The Sustaining Effects Study, Final Report. Santa Monica, CA: System Development Corporation.

Carter, L. F.
1980 The Sustaining Effects Study, An Interim Report. Washington, DC: U.S. Department of Education, Office of Program Evaluation.

Coleman, J., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld and R. L. York
1966 Equality of Educational Opportunity. Washington, DC: U.S. Government Printing Office.

Cooley, W. W.
1978 "Explanatory Observational Studies." Educational Researcher 7: 9-15.

Cooley, W. W. and G. Leinhardt
1980 "The Instructional Dimensions Study." Education Evaluation and Policy Analysis. 2(1).

Cooper, H. M.
1986 "Chapter 1 Programs Reduce Student-to-Instructor Ratios: But Do Reduced Ratios Affect Achievement?" Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

de Kanter, A., A. L. Ginsburg and A. M. Milne
1986 "Parent Involvement Strategies: A New Emphasis on Traditional Parent Roles." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Freeman, D. J., T. M. Kuhs, A. C. Porter, R. E. Floden, W. H. Schmidt, and J. R. Scheville
1983 "Do Textbooks and Tests Define a National Curriculum in Elementary Mathematics?" Elementary School Journal. 83: 501-513.

Gabriel, R. M., B. L. Anderson, G. Benson, R. Hill, J. Pfannensteil and R. M. Stonehill
1985 The Sustained Achievement of Chapter 1 Students. Washington, DC: U.S. Department of Education.

Griffin, G.
1986 "Chapter 1 and the Regular School: Staff Development." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Griswold, P. A., K. J. Cotton and J. B. Hansen
1986 Effective Compensatory Education Sourcebook. Washington, DC: U.S. Government Printing Office.

Hayes, D. P. and J. Grether
1969 "The School Year and Vacations: When Do Students Learn?" Revision of a paper presented at the Eastern Sociological Association Convention, New York, April. Mimeographed.

Hemenway, J., M. Wang, C. Kenoyer, R. Hoepfner, M. Bear, and G. Smith
1981 Report 9: The Measures and Variables in the Sustaining Effects Study. Santa Monica, CA: System Development Corporation.

Heyns, B.
1978 Summer Learning and the Effects of Schooling. New York: Academic Press.

Heyns, B.
1986 "Summer Programs and Compensatory Education: The Future of an Idea." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Holley, F. M.
1986 "Program and Staff Structure: Reactions from a Quarter-Century Worker in Compensatory Education." Paper prepared for the Conference on the Effects of Alternative Designs in Education. Washington, DC: Office of Educational Research and Improvement.

Karweit, N.
1983 Time on Task: A Research Review (Report No. 332). Baltimore, MD: Center for the Social Organization of Schools.

Keesling, J. W.
1984 Differences Between Fall-to-Spring and Annual Gains in Evaluation of Chapter 1 Programs. McLean, VA: Advanced Technology, Inc.

Kennedy, M. M., R. K. Jung and M. E. Orland
1986 Poverty, Achievement and the Distribution of Compensatory Education Services. Washington, DC: Office of Educational Research and Improvement.

Kirshstein, R. J.
1986 Chapter 1 Students in the Montgomery County Public Schools: A Description and Follow-up. Rockville, MD: Montgomery County Public Schools.

Klibanoff, L. A. and S. A. Haggart
1981 Report 8: Summer Growth and the Effectiveness of Summer School. Santa Monica, CA: System Development Corporation.

Koretz, D.
1986 Trends in Education Achievement. Washington, DC: Congressional Budget Office.

LaPointe, A. E.
1984 Statement to the Subcommittee on Elementary, Secondary and Vocational Education, Committee on Education and Labor, U.S. House of Representatives. January 31.

Linn, R. L.
1982 "The Validity of the Title I Evaluation and Reporting System," in E. R. Reisner, M. C. Alkin, R. S. Boruch, R. L. Linn, and J. Millman, Assessment of the Title I Evaluation and Reporting System. Washington, DC: U.S. Department of Education.

Linn, R. L.
1979 "Validity of Inferences Based on the Proposed Title I Evaluation Models." Educational Evaluation and Policy Analysis 1: 23-32.

McLaughlin, M. W. and P. M. Shields
1986 "Involving Parents in the Schools: Lessons for Policy." Paper prepared for the Conference on the Effects of Alternative Designs in Education. Washington, DC: Office of Educational Research and Improvement.

McPartland, J. and R. L. York
1967 "Further Analysis of Equality of Education Opportunity Survey." Pp. 35-142 in U.S. Commission on Civil Rights, Racial Isolation in the Public Schools 2. Washington, DC: U.S. Government Printing Office.

Mullins, S. P. and A. A. Summers
1983 "Is More Better? The Effectiveness of Spending on Compensatory Education." Phi Delta Kappan 64: 339-347.

Myers, D. E.
1986 "An Analysis of the Impact of Title I on Reading and Math Achievement of Elementary School Aged Children." Washington, DC: Office of Educational Research and Improvement.

National Assessment of Educational Progress
1985 The Reading Report Card, Progress Toward Excellence in Our School, Trends in Reading Over Four National Assessments, 1971-1984. Princeton, NJ: Educational Testing Service.

National Institute of Education
1977 The Effects of Services on Student Development. Washington, DC: U.S. Department of Health, Education and Welfare.

National Institute of Education
1978 Compensatory Education Study. Final Report to Congress from the National Institute of Education. Washington, DC: author.

National Institute of Education
1978 Perspectives on the Instructional Dimensions Study. Washington, DC: U.S. Department of Health, Education and Welfare.

Pfannensteil, J.
1986 Analysis of Categorical Program Participation and Long-Term Effects. Overland Park, KS: Research and Training Associates, Inc.

Plato, K., D. Deck, G. Ensign, D. MacQuairie, and P. Neill-Carlton.
1986 A Study of Categorical Program Participation of Chapter 1 Students. Olympia, WA: Superintendent of Public Instruction.

Rich, D.
1986 "The Parent Gap in Compensatory Education." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Riddle, W.
1984 Achievement Score Trends and Federal Involvement in Elementary and Secondary Education: An Exploration of Their Relationships (Report 84-627EPW). Washington, DC: Library of Congress, Congressional Research Service.

Romberg, T. A.
1986 "Mathematics for Compensatory School Programs." Paper prepared for the Conference on the Effects of Alternative Designs in Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Smith, A.
1985 The Impact of Headstart on Children, Families, and Communities. Final Report of Headstart Synthesis Project (HHS Report (OHDS) 85-31193). Washington, DC: U.S. Department of Health and Human Services.

Smith, M. S.
1986 "Selecting Students and Services for Chapter 1." A reactor's paper prepared for the Conference on the Effects of Alternative Designs for Compensatory Education. Washington, DC: Office of Educational Research and Improvement.

Walberg, H. J. and W. C. Frederick
1983 "Instructional Time and Learning," in H. Mitzeo, ed. Encyclopedia of Educational Research, Volume 2. New York: Free Press.

Wang, M., M. Bear, J. Conklin, and R. Hoepfner
1981 Report 10: Compensatory Services and Educational Development in the School Year. Santa Monica, CA: System Development Corporation.

White, K. R.
1982 "The Relationship Between Socioeconomic Status and Academic Achievement." Psychological Bulletin 9: 461-481.

Williams, B. J., P. A. Richmond and B. J. Mason
1986 Effects of Alternative Designs in Compensatory Education - Designs for Compensatory Education: Conference Proceedings and Papers. Washington, DC: Office of Educational Research and Improvement.

Wolf, A.
1977 "The Relationship Between Poverty and Achievement." Compensatory Education Study. Washington, DC: National Institute of Education.

Wood, C. T.
1984 Uses of Data from State Title I/Chapter 1: Evaluation and Performance Reports for Assessing the Impact of Compensatory Education. Mountain View, CA: SRA Technologies.

APPENDIX A

Statute for the National Assessment of
Chapter 1

NATIONAL ASSESSMENT OF COMPENSATORY EDUCATION ASSISTED UNDER THIS CHAPTER

"Sec. 559 (a) The Secretary shall conduct a national assessment of compensatory education assisted under this chapter, through independent studies and analysis by the National Institute of Education. The assessment shall include descriptions and assessments of the impact of (1) services delivered, (2) recipients of services, (3) background and training of teachers and staff, (4) allocation of funds (to school sites), (5) coordination with other programs, (6) effectiveness of programs on student's basic and higher order academic skills, school attendance, and future education, and (7) a national profile of the way in which local educational agencies implement activities described under section 556(b). The National Institute of Education shall consult with the Committee on Labor and Human Resources of the Senate and the Committee on Education and Labor of the House of Representatives in the design and implementation of the assessment required by this section. The National Institute of Education shall report to Congress the preliminary results of the assessment required by this section in January and July of 1986, and a final report shall be prepared and submitted to the Congress not later than January 1, 1987.

"(b) Notwithstanding any other provision of law or regulation, such reports shall not be subject to any review outside of the Department of Education before their transmittal to the Congress, but the President and the Secretary may make such additional recommendations to the Congress with respect to the assessment as they deem appropriate.

APPENDIX B

Administrative Status of the National Assessment

ADMINISTRATIVE STATUS REPORT

The National Assessment of Chapter 1 was legislated as part of the Technical Amendments to the Education Consolidation and Improvement Act of 1981. Enacted in December 1983, these amendments required the National Institute of Education (NIE) to provide Congress with a final report by January 1987, just three years later. The major milestones for this National Assessment, either accomplished or projected, are listed in Table 1. The remainder of this administrative status report reviews each aspect of the administration of the National Assessment of Chapter 1.

Milestones

Following passage of the legislation, NIE hired a Study Director to design the National Assessment of Chapter 1 and to oversee its implementation. The Study Director joined NIE in April 1984.

During the summer of 1984, the Study Director discussed the study's purposes and Congress' information needs with several Congressional staff members; with Department of Education staff, both within the Chapter 1 program administration and in the Department's Office of Planning, Budgeting and Evaluation; with members of the Office of Management and Budget, the Congressional Research Service, the Congressional Budget Office, and the General Accounting Office; with many educational associations and interest groups within the Washington area which were known to have an interest in the Chapter 1 program and its future; and with a variety of educational researchers and program evaluators.

On the basis of these discussions, NIE developed a plan for the National Assessment of Chapter 1. The plan was reviewed by Department of Education officials in late summer and in October it was presented to Congressional staff. Two briefings

Table 1

Milestones for the National Assessment of the Chapter 1 Program

December 1983	Congress passes legislation requiring the National Institute of Education to conduct a National Assessment of the Chapter 1 program.
April 1984	National Institute of Education hires a Director to oversee the National Assessment.
October 1984	National Institute of Education presents a Study Plan to the Congress.
December 1984	National Institute of Education completes hiring a Study Team to implement the Study Plan.
May - September 1985	National Institute of Education procures a series of independent studies for the National Assessment.
October 1985	National Institute of Education is replaced by the Office of Research within the Office of Educational Research and Improvement.
January 1986	Office of Research produces its first Interim Report, as required by the Congress, for the National Assessment of Chapter 1.
July 1986	The Office of Research is scheduled to produce its Second Interim Report for the National Assessment of Chapter 1.
January 1987	The Office of Research is scheduled to produce its Final Report for the National Assessment of the Chapter 1 program.

were held, one for Senate staff and one for House staff. Following these briefings, further changes were made in the plan.

In November 1984, the final plan was submitted to Congress and NIE began in earnest to implement it. Concurrently, NIE began forming a Study Team to implement the plan. Qualified researchers both within the Department and outside it were solicited. A few Department staff began in the summer of 1984, but researchers from outside the Department did not join the Study Team until December.

The Study Plan outlined a number of separate investigative components which, taken together, would provide information regarding the full range of questions and issues that had been raised during the preceding summer. The first stages of implementation of the Study Plan consisted primarily of contracting assistance in carrying out a number of these components. NIE chose to contract out portions of the work in part because the level of effort involved in doing these projects did not make in-house work feasible, and in part because contracted studies assure a level of independence often necessary to give the overall study credibility.

Requests for proposals for these studies were prepared throughout early 1985, advertised through the spring of the year, and contracts for the projects were awarded throughout the summer. The full list of procured studies appears in Appendix C.

In October 1986, a year after NIE presented its Study Plan to Congress, the Secretary of Education reorganized the Office of Educational Research and Improvement (OERI) in such a way as to integrate NIE's components into the rest of the office. Under the reorganization, the Chapter 1 Study Team was located in the Office of Research (OR), one of the five components within the Office for Educational Research and Improvement. The Office of Research contains the research functions that had been previously placed in NIE, and consequently is the closest approximation to NIE that now exists.

This report constitutes the second of three reports which will be delivered to Congress as part of this National Assessment of Chapter 1. The Final Report will describe findings from all contracted studies, summarizing virtually every aspect of the current operation of Chapter 1 programs.

Budget

The budget for the National Assessment has proved to be one of its most complicated and problematic features. This has occurred for three reasons. First, the study was expected to be funded from three sources, rather than one, thus requiring three separate budget lines rather than one. Second, one of those sources, the Chapter 1 budget, is forward-funded. This means that its budget does not normally become available until three-quarters of the way through the fiscal year. For fiscal year 1985, Chapter 1 funds could not contribute to the study until well into the fiscal year. For fiscal year 1986, the Department asked Congress to make a special provision in the Chapter 1 budget which would enable Chapter 1 funds being used for this National Assessment to become available earlier. The special legislation resulting from that request greatly facilitated the progress of the National Assessment. Finally, the third source of funds, the Secretary's Discretionary Fund, was impounded by the Federal District court in Chicago, and therefore was not available at the time or in the amount that the Chapter 1 study had anticipated. As a result of these budgetary difficulties, the Study Team solicited funds from programs within the Department of Education other than those specified by the Congress. NIE signed agreements with both the Office of Bilingual Education and the Office of Special Education such that these offices agreed to support studies of topics of mutual interest. However, also as a result of these budgetary difficulties, many of the procurements were postponed because funds were not available when they were originally anticipated. Finally,

virtually every project had to be incrementally funded so that those funds that were available at first could be used to start as many projects as possible.

Table 2 summarizes the contributions made by each funding source to date.

Table 2

Contributions to the National Assessment of Chapter 1
(in thousands)

Funding Source	FY '83	FY '84	FY '85	FY '86	Total
Chapter 1	$376	$400	$1,500	$1,100	$3,376
NIE		300	450	1,481	2,231
Chapter 2 (Secretary's Fund)			800	450	1,250
Bilingual Education			350	75	425
Special Education			200		200
TOTAL	$376	$700	$3,300	$3,106	$7,482

No FY '87 funds have been appropriated for further analyses of the National Assessment's data.

Current Status

The current status of the National Assessment of Chapter 1 is as follows:

- Most funded projects are nearing completion. Data were collected from school districts during the 1985-86 school year. Data are being analyzed and final reports are being prepared.
- The Final Report to Congress is being prepared.

Despite difficulties in awarding contracts and getting contracted studies underway, OERI still plans to complete the National Assessment on schedule. However, the timeline is such that many interesting analyses will not be completed. The staff of the

National Assessment plans to present a series of separate special-topic reports in early 1987 to enhance the findings presented in the main report. Some topics being considered for special reports are: services provided to private school students by Chapter 1, the provision of Chapter 1 services to LEP students, the relationship between Chapter 1 and special education programs, and the role of States in providing compensatory education services.

APPENDIX C

Studies Commissioned by the National
Assessment of Chapter 1

A Study of Local Implementation of ECIA Chapter 1

Contractor: Research and Evaluation Associates, Inc., Washington, DC
Subcontractor: Westat, Inc., Rockville, Maryland
Project Officer: Ron Anson

This survey of 2,000 school districts included questions about all the major provisions in the legislation:

- parent involvement
- program evaluation
- needs assessment
- selection of schools and students
- services to private school students
- program design and resource allocation
- administration and record-keeping

The survey sample was designed to partially overlap with the sample of school districts which participated in the 1980-81 District Practices Survey, thus permitting cross-time comparison in these districts.

In addition to the mailed questionnaire to 2,000 school districts, the survey included in-depth telephone interviews with 200 of the districts which responded to the mailed questionnaire.

A Survey of Chapter 1 Schools and Teachers

Contractor: Westat, Inc., Rockville, Maryland
Subcontractor: RMC Research Corp., Hampton, New Hampshire
Project Officer: Gil Garcia

These researchers surveyed by telephone staff members in roughly 1,300 schools across the country. Principals of Chapter 1 schools were asked about the characteristics of their schools and about their Chapter 1 program configurations and

Chapter 1 teachers were asked about their education and experience and about the services they actually provide to students. For comparison purposes, regular classroom teachers and teachers of other categorical programs were also sampled and asked analogous questions.

The survey included:

- Elementary Schools with Chapter 1 Programs
  - Some with high concentrations of poverty.
  - Some with high concentrations of limited-English proficient Students.
- Elementary Schools Without Chapter 1 Programs
  - Some with other kinds of compensatory education.
  - Some with no compensatory education students but with disadvantaged students.
  - Some with no compensatory education and with very few disadvantaged students.
- Private Elementary Schools with Chapter 1 Programs
- Middle Schools and Secondary Schools
  - Some with Chapter 1 Programs
  - Some with other compensatory education programs

A CES Fast Response Survey (FRS) of Chapter 1 Oversight

Agency: Center for Educational Statistics
Project Officer: Marty Orland

This survey of 700 school districts contained questions about local experiences with State program monitoring, state audits, Federal management reviews and Federal audits. Questions were asked regarding the number of experiences districts have had with each type of oversight, the content covered by the oversight review and the nature of changes in practices, if any, that resulted.

A CES Fast Response Survey (FRS) of Private School Students' Participation in Chapter 1

Agency: Center for Educational Statistics
Project Officer: Richard Jung

This survey of 900 school districts provided national information about the number of nonpublic school students participating in the Chapter 1 program as well as the location and time (e.g., before or after school) of such services. Data on these and related issues were collected for the 1986-87 school year and, retrospectively, for the 1984-85 school year, in order to compare levels and types of Chapter 1 services for nonpublic school students before and after the Aguilar v. Felton Supreme Court decision. In this decision, the court ruled that the provision of Chapter 1 instructional services in sectarian schools was unconstitutional.

A Study of the Costs of Special Education Services, Amended to Include Costs of Chapter 1 Services

Contractor: DRC, Inc., Washington, DC
Project Officer: Marty Orland

This study, funded initially by the Education Department's Office of Special Education, was designed to determine the resources (and their costs) used to provide special education services under a variety of specific program arrangements. NIE amended the contract so that the investigators were able to document the resources used to provide services under a variety of Chapter 1 program arrangements as well. The Chapter 1 program options include elementary in-class programs, elementary pull-out programs, secondary programs, after school programs and so forth. The study was conducted in a nationally representative sample of 60 school districts.

The National Assessment of Educational Progress (NAEP)

Contractor: Educational Testing Service, Princeton, New Jersey
Subcontractor: Westat, Inc., Rockville, Maryland

Though the primary purpose of the NAEP is to document student achievement, it also includes data on students' school and home experiences, teacher and classroom characteristics and school characteristics. The National Assessment of Chapter 1 designed an amendment to the NAEP grant to conduct special analyses regarding Chapter 1 students and Chapter 1 schools. Data will not be available in time for our final report.

The Title I Evaluation and Reporting System (TIERS)

This reporting system, developed in response to Title I regulation, continues to exist in a reduced form under Chapter 1. Local education agencies report to State agencies who in turn report to the U.S. Education Department. These reports include data regarding the number of students served by grade level and service provided.

CASE STUDIES

A Study of Administration

Contractor: Abt Associates, Inc., Cambridge, MA
Subcontractors: Education Commission of the States, Denver, CO
Policy Studies Associates; Washington, DC
COSMOS, Washington, DC
Project Officer: Marty Orland

Investigators documented administrative practices in both State and local educational agencies during the 1985-86 school year and any major changes that occurred since Title I. Topics included:

- At the State level:
  - Monitoring and enforcement
  - Technical assistance
  - Application approval
  - Policies in areas where the Federal law has changed, such as parent involvement, comparability, and evaluation.
- At the local level:
  - Needs assessment and evaluation
  - Program design decision-making
  - Funds allocation policies and practices
  - Parent involvement
  - Application and reporting activities

The study entailed visits to 20 States, with return visits to nine of them. Then, in each of the nine States, three school districts were visited.

A Study of School District Program Design Decision-Making

Contractor: SRI International, Menlo Park, California
Subcontractor: Policy Studies Associates, Washington, DC
Project Officer: Ron Anson

Investigators determined how districts and schools make program design decisions and compared districts and schools that changed their approaches to those that remained constant. The study had two goals:

- To gain a better understanding of why districts and schools change or maintain key features of their Chapter 1 programs. The study examined the influences of:

  - Legislative change from Title I to Chapter 1
  - Shifts in State or local policies
  - Changes in budget contexts
  - Program design preferences of State or district administrators and teachers
  - Apprehension about Federal audits
  - Institutionalization of the Chapter 1 program
  - Conviction that the program is successful and working well

- To examine decisions to adopt or forego particular program design features of current public interest. Examples of such features are:

  - Programs in secondary schools
  - In class program designs
  - Reliance on aides vs. teachers
  - Schoolwide projects
  - Changes in the intensity or grade levels of services
  - Parent involvement
  - The use of computers
  - Emphasis on higher order skills

The study was conducted in 20 school districts and 60 schools.

<u>A Study of How Districts Allocate Resources Among Schools</u>

Contractor:	Educational Testing Service, Princeton, New Jersey
Subcontractors:	Gaffney, Anspach, Schember, Klimaski and Marks, P.C., Washington, DC
	Decision Resources, Inc., Washington, DC
Project Officer:	Marty Orland

Investigators examined both the decision-making processes used to allocate resources among schools and the resource distributions--the number of teachers, aides, or computers per child or per grade level that resulted from those decisions.

Investigators:

- Described the influence of State and Federal laws on local decisions.

- Determined whether different decision-making strategies yield different patterns of resource allocation.

- Showed the effect of different resource allocation strategies on economically- and educationally-disadvantaged schools and students.

- Determined the effect of multiple-needs students, and of multiple Federal and state programs on Chapter 1 resource allocation patterns.

- Described changes in resource allocation from Title I to Chapter 1.

Twenty districts in eight States were visited. Investigators interviewed both district and school staff regarding their decision making, and documented actual resources that resulted from those decisions.

A Study of Targeting Practices Used in the Chapter 1 Program

Contractor: SRA Technologies, Inc., Mountain View, California
Subcontractor: Northwest Regional Educational Laboratory, Portland, OR
Project Officer: Dick Jung

Investigators examined the net effects of Chapter 1 school and student selection procedures on the characteristics of the students served in the program. The study analyzed data on student poverty level, achievement status, grade point average, attendance rates, grade retention patterns, and participation in other categorical programs for Chapter 1 and non-Chapter 1 students in 30 districts. The five major questions addressed were:

- How do districts determine which schools and students receive program services?

- What rationale(s) underlie district policies and practices for selecting project schools and participants?

- How do Chapter 1 schools and students differ from non-Chapter 1 schools and students?

- Are different types of students served under Chapter 1 than were served under Title I?

- What are the effects of varying school and student selection practices on the characteristics of students served in the program?

Thirty districts were selected so that they were diverse with respect to size, urbanicity, region, poverty level, concentration of limited-English-proficient students, and the presence or absence of non-Federal compensatory education programs.

A Study of the Whole-Day Instructional Experiences of Chapter 1 Students

Contractor: Far West Educational Laboratory, San Francisco, CA
Project Officer: Randy Demaline/Gil Garcia

Investigators described the actual configuration of services that Chapter 1 students receive over the course of a school day and over the course of a school week. They determined:

- Students' exposure to various instructional topics.
- How services are coordinated across service providers.
- The quality of instruction provided.
- The services provided in the regular classrooms while Chapter 1 students are pulled out.
- What teachers and students perceive the role and purpose of Chapter 1 to be.

The study included 24 schools distributed over six geographic regions. The schools encompassed elementary, middle, and secondary levels, and some were private schools. Within each school eight students were followed for a day and two were followed for a week. The students will vary in grade level and achievement levels, and in the configuration of services provided to them.

Effects of Alternative Designs in Compensatory Education

Contractor: Research and Evaluation Associates, Washington, DC
Project Officer: Randy Demaline/Ron Anson

A number of independent scholars were asked to review research on the effectiveness of various program design features used in compensatory education. Features included the following:

- Staffing patterns
- Service configurations
- Relationship between Chapter 1 and the regular program
- Curriculum
- Overall compensatory education strategies

Once these research summaries were completed, other researchers were asked to critique them. All summaries and critiques were then reviewed by a panel of educators, and their findings have been bound together in a summary document.

Alternative Strategies in Compensatory Education

Contractor: Research and Evaluation Associates, Washington, DC
Project Officer: Ron Anson

A number of prominent authors and practitioners were asked to prepare papers on issues that policy makers may consider in the upcoming reauthorization of Chapter 1, ECIA. The issues related to:

- The Federal role
- Access and accountability
- Program implementation

The papers were reviewed and critiqued by the authors and other researchers and practitioners. After responding to the critiques, final versions of the papers will be bound in a summary document.

OTHER STUDIES

Analysis of School District and State Education Agency Records

Contractors: Seven contracts have been awarded to school districts or State agencies.
Project Officer: Gil Garcia

State and local agencies analyzed their data bases on Chapter 1 students to answer questions regarding the coordination of Chapter 1 with other programs and regarding the effectiveness of Chapter 1 services. Two categories of studies were funded:

- Investigations of the patterns of categorical services Chapter 1 students receive over several years; and
- Investigations of the long-term educational accomplishments of compensatory education program students.

Each State or local agency conducted analyses that were appropriate to the particular data bases it had available.

Data Analysis and Technical Support

Contractor: DRC, Inc., Washington, DC
Subcontractor: Policy Studies Associates, Washington, DC
Project Officer: Bea Birman

This contract serves a number of purposes for the National Assessment of Chapter 1. The contractor is responsible for:

- Creating a data library
- Conducting computer analyses of large data bases (e.g., Census, Sustaining Effects Study)
- Conducting literature reviews and issue analyses
- Coordinating data collection and analyses across all of the other procured studies

These activities will provide information that the National Assessment will use in writing its three Congressionally mandated reports. This contract has already produced

Chapter 1 state profiles, an overview of State compensatory education programs for the National Assessment of Chapter 1, and studies of Federal administration of the Chapter 1 program.

APPENDIX D

Description of Data Sources Used in This Report

DESCRIPTION OF DATA SOURCES USED IN THIS REPORT

This appendix describes the data sources used in the report and assesses their technical quality and applicability. The review of each data source presents background information and describes relevant technical aspects, including sampling procedures, student and program measures, and measures of outcomes or impacts. Each review then evaluates the strengths and weaknesses of the data source.

The following data sources are reviewed:

- The Instructional Dimensions Study (IDS);
- The Sustaining Effects Study (SES);
- The Title I/Chapter 1 Evaluation and Reporting System (TIERS);
- The National Assessment of Educational Progress (NAEP);
- The CBO Report on Trends in Educational Achievement;
- The Sustaining Achievement Study (SAS); and
- The study entitled Summer Learning and the Effects of Schooling.

NIE Instructional Dimensions Study

I. Background and Study Characteristics

A. Overview of the Study

The IDS was one of several studies conducted by the National Institute of Education (NIE) as part of the mandate of the 1974 Educational Amendments to study compensatory education supported by ESEA Title I and by various States. Based on data collected in school year 1976-77 with follow-up data collected in the fall of 1977, the IDS sought to identify relationships between a variety of classroom structures and processes and program effectiveness. The data for the study were collected from 400 purposively selected first- and third-grade classrooms. The IDS assessed program impacts on reading and mathematics achievement scores.

One major purpose of the study was to examine the differential effects of pullout and in class designs on the achievement of students who received compensatory education. The second major focus of the IDS was assessing the impact of several instructional practices and characteristics--including instructional time, individualization, teacher-student interactions, classroom climate and motivation, and teacher training and experience--on program outcomes.

B. Sampling Procedures

NIE's approach to sampling for the IDS was to select classrooms that evidenced specific instructional dimensions, without specifically seeking to produce nationally representative results. At the same time, NIE aimed at selecting a wide variety of classroom settings comparable to the diversity of the Title I program nationwide in 1976. Therefore, classrooms were selected that varied on student background factors, such as family income levels and ethnicity. In addition, classrooms were selected from school districts that varied in urban status, size, geographic location, and percent of Title I enrollment.

NIE sought nominations from SEA and LEA officials, curriculum publishers, and lists of exemplary programs. Nominated districts and schools were then visited prior to final sample selection to ensure that practices of interest had actually been implemented and that those practices had been in place for at least one year. The final sample was made up of 400 classrooms (200 first grade and 200 third grade) in 100 schools from 14 districts.

C. Student, Program, and Outcome Measures

There is no evidence in the report on findings from the IDS (NIE, 1977) that background data on students were collected or used in the analysis.

Extensive data were collected on instructional practices employed in the 400 classrooms studied by NIE. These data included (1) interviews with principals and

teachers to assess teacher background and experience, (2) curricular analysis to judge the overlap between curriculum and the content of the outcome measure (the Comprehensive Test of Basic Skills--CTBS), and (3) videotapes of each classroom to provide detailed data on classroom characteristics such as teacher-student interactions and time on task. The IDS also collected information on costs associated with each reading and mathematics program.

The study used the CTBS, level B, form S to measure achievement of first graders and CTBS level I, forms S (fall) and T (spring) for third-grade measurement. Students' attitude toward school was also measured in fall and spring. However, little analysis of this measure was reported except to relate in a footnote that fall measures of attitudes were high and remained high at the time of spring assessment.

II. Strengths and Weaknesses

The IDS has several strengths that make it relevant for any assessment of Chapter 1's effectiveness. First, it assesses the effectiveness of compensatory education practices. Some of the other studies reviewed in this technical appendix do not assess program impacts, or they focus on broader populations than just the educationally disadvantaged. Second, the IDS collected the most comprehensive data on compensatory education practices for a comparatively large sample of classrooms. The study went beyond mean effects to examine the differential effects of pullout and in-class designs--the most common Chapter 1 instructional arrangements--and to evaluate finer grained classroom processes and characteristics such as time spent on specific learning tasks, the match of instruction to student needs, incentives provided for learning, and teacher training and experience. Third, unlike TIERS, which contains data on effectiveness from whatever tests local school districts chose to use, the IDS used the same measure of achievement for all 400 classrooms. In addition, the study assessed impact on achievement gains of overlap between items in the spring test and

instructional content. ("For all groups of students, overlap between curriculum content and test items has a strong positive relationship to achievement" [p.32]).

At the same time, the IDS has several weaknesses that limit the study's usefulness. First, the data are 10 years old. While data on the efficacy of pullout and in-class designs and the impacts of classroom processes may still be relevant, the IDS obviously cannot assess whether changes in Title I/Chapter 1 since 1976 have influenced the effectiveness of the program. For example, other program models such as the replacement design and extended pullout programs were not specified in the Title I legal framework until 1978. The IDS tells us nothing about these approaches. Moveover, compensatory education is a more mature program, and staff have 10 years more experience on how to teach the educationally disadvantaged. Findings from the IDS and other studies--such as those about time on task and the possible inappropriateness of pullout designs for young children--have influenced the design of Chapter 1 programs. Ten-year-old data can tell us nothing of the impact of these changes.

The sample of the IDS also limits its usefulness. The authors of this study acknowledge that their aim was not to collect nationally representative effectiveness data. Their sample was chosen purposively to ensure that practices of interest were included. Thus, their conclusions cannot be generalized beyond schools and programs in their sample.

Finally, although the IDS apparently collected extensive details on classroom process, these data seem to have been underutilized. For example, the study reports larger average gains in raw scores, grade equivalents, and percentiles for those first graders who were mainstreamed compared to those who were in pullout programs. However, significance tests are not reported, and there is little assessment of the educational importance of these gains. With the wealth of data collected, it would

seem to have been possible to explore reasons why, for example, a pullout design was found to be less effective for first graders but more effective for third-grade mathematics programs.

The Sustaining Effects Study

I. Background and Study Characteristics

A. Overview of the Study

The SES was a Federally mandated investigation of compensatory education, including Title I and State and local programs. Conducted by System Development Corporation (SDC), the study examined the "nature, quantity, and environment" of compensatory education and "its sustained effects" (Hinckley, 1979, p. xvi). Funded at more than $20 million, the SES conducted a series of substudies including a longitudinal study, a cost/effectiveness study, a study of summer programs, a study of successful sites, and a participants' study. (Findings from some of these substudies were never published, however, because funding ran out.) The first year (1975-76) of the study was used for planning, instrument development, and sample selection. Data were then collected in the fall of school year 1976-77 and in the next three school years.

The SES addressed two major policy issues:

- Who receives compensatory education?
- How effective is compensatory education?

In addition, the study examined several subsidiary issues:

- What is compensatory education?
- What is the nature of elementary school children's home environment, and how is home environment related to school environment?
- What happens to students' achievement when compensatory education is discontinued?
- What is the optimum duration and time for receiving compensatory education services?

- What happens to achievement over the summer, and are summer school programs effective?

B. Sampling Procedures

The SES data and analyses were based on three samples: the Representative Sample (drawn to be representative of students in the nation's schools), the Comparative Sample (composed of 29 schools identified in high poverty areas but not receiving compensatory education resources), and the Nominated Sample (composed of 43 Title I schools nominated by Federal, State, and local officials as having particularly effective compensatory education practices).

The Representative Sample is the main source of SES data used in this report. It included 243 schools that were selected in the first stage of the sampling. These schools were randomly selected to represent schools stratified by region, size of district, and district poverty status. Schools from high-poverty districts were oversampled. Based on enrollment lists for fall 1976, the second stage sample was a systematic random sample of children enrolled in the schools. Certain subsets of schools and students were excluded from the sample: for example, schools that served mainly handicapped students and students in bilingual or English-as-a-Second-Language programs. Because the study was never fully funded, the sample in follow-up years was reduced. The Representative Sample was reduced by 60 percent, and, overall, 70,000 of the 120,000 students who participated in the first year of the study were included in the follow-up data collection.

C. Student, Program, and Outcome Measures

The SES collected massive amounts of data on students, teachers, schools, compensatory education programs, and school districts. Student outcomes in reading and mathematics were measured with the CTBS. In addition, students were administered a functional literacy test and an attitudinal survey of their views of school and of reading and mathematics.

For each year of data collection, teachers completed a questionnaire with sections on characteristics of the school and the teacher's background and details about instructional programs including whether grouping was used and what and how instructional materials were employed. Principals completed questionnaires on their backgrounds and school characteristics. The superintendent and business officer provided data on the district and on the Title I program. Homeroom teachers provided background information about students and tracked students' program participation and attendance. Finally, a school coordinator--paid by the study--maintained a Compensatory Education Roster, which classified students according to the type of compensatory education services they received.

II. Strengths and Weaknesses

The strengths and weaknesses of the SES can be approached from two perspectives: criticisms of the reports from the substudies and strengths and weaknesses of the data base, which is available for reanalysis. After a brief review of criticisms of the reports, the strengths and weaknesses of the data base will be examined.

The SES draft and final reports have been the subject of considerable criticism. The report of the majority of a special panel that was convened to review the SES (Hanushek, Breneman, & Hauser, 1979) made the following general criticisms of the substudy reports:

- The technical quality and exposition of the reports on the whole are below prevailing standards and in some cases are unacceptable.
- Statistical analyses have serious flaws.
- The structure and reporting of analyses limit or even preclude their usefulness for evaluating or developing future compensatory education policies.

Other critics were less harsh. Linn (1979), for example, who was also a member of the review committee, did not concur with the majority report and submitted a separate memorandum with his comments. He concluded in part that "although...the reports have flaws, they also contain some important and potentially enlightening information....[E]ven the two weaker reports [#3 and #4] provide considerable information about the characteristics of participants in [compensatory education], especially Title I" (p. 2).

Even severe critics admit that the SES data base has strengths. The majority review report praises the first year sample and notes that those data are useful for describing who participates in compensatory education. That report also lauds the development, administration, and scaling of achievement tests used in the SES. Linn (1979) points out that administration of tests by the contractor is a particular strength of the study and that "the norming and vertical scaling of the achievement tests are about as good as could be hoped for" (p. 1).

At the same time, critics have raised important problems with the SES data that must be considered when either examining substudy results or reanalyzing the SES data. The following are the most serious problems with the data base:

- The attrition of schools and students from the longitudinal sample, especially as it limits the generalizability of data to the national population; and
- The absence or poor quality of data for measuring variables, other than student achievement, needed to assess the effectiveness of compensatory education, e.g., student background, student status in compensatory programs, and the characteristics of such programs.

Reanalyses of the SES data base need to adjust for these shortcomings, where possible.

The Title I/Chapter 1 Evaluation and Reporting System (TIERS)

I. Background and Study Characteristics

A. Overview of the Study

The TIERS was mandated by Congress as part of the Education Amendments of 1974. As developed by the U.S. Office of Education, TIERS had several features including models for evaluating Title I programs, technical assistance centers (TACs) to aid in the development and implementation of these models, State and local reporting requirements, and dissemination of exemplary Title I practices to states and local school districts.

In 1981, ECIA Chapter 1 repealed the authorization for TIERS as part of an effort to reduce Federally imposed burden, paperwork, and supervision. Mandatory data collection and reporting of evaluation data occurred in the last three years of Title I (1979-80, 1980-81, and 1981-82). Since the enactment of ECIA, collection, analysis, and reporting based on TIERS have been voluntary. The most recent published analyses of TIERS data are based on data from the 1983-84 school year (Carpenter & Hopper, 1985).

B. Sampling Procedures

Unlike some studies discussed in this appendix, TIERS does not depend on sampling techniques to estimate achievement gains attributable to Title I. (However, states can sample districts for reporting purposes.) Instead, it is a reporting system that aggregates data on gains from the project level to the district and then to the state level. States analyze and summarize district data to prepare reports to ED. These data--aggregated at various levels--are intended to assess how much additional gains in basic achievement levels were the result of Title I projects.

C. Student, Program, and Outcome Measures

TIERS collects demographic data on program participants and certain descriptive information (e.g., the number of students participating in Title I, the amount of parent involvement in Title I projects, characteristics of Title I staff, and services that Title I provides to students.) TIERS also reports gains in achievement in the basic skills of reading, mathematics, and language arts. Selection of achievement tests is left to the discretion of the States and local school districts to help ensure that the tests used are closely matched with the goals and content of local projects. Changes in achievement are reported in Normal Curve Equivalents (NCEs). Each NCE value has a corresponding percentile rank so that, for example, "an NCE of 10 corresponds to a percentile rank of 23" and is interpreted as "23 percent of the norm group scores below an NCE of 10" (Linn, 1982, p.9).

TIERS provides three basic models (Models A, B, and C) for analyzing Title I achievement gains. Since almost every district uses Model A (the norm-referenced model), discussion of the strengths and weaknesses of the TIERS data will be restricted to that model. As its name implies, Model A uses test publishers' norms to estimate what students would have achieved if they had not received Title I services. Model A "simply assumed that, in the absence of participation in a Title I project, participants would maintain, on the average, a constant position relative to the norm" (Linn, 1982, p.10).

II. Strengths and Weaknesses

TIERS represents a useful attempt to provide school districts, states, and the federal government with data on the impact of the Title I program. Unfortunately, as critics have shown, there are several reasons for viewing TIERS data skeptically.

Statistical and other technical shortcomings include the following:

- Regression effects: Although Model A was designed to reduce bias resulting from regression effects, even the developers of the evaluation models (RMC Research Corporation) admit that the regression effect is not completely eliminated and results in overestimates of about 1 NCE.

- Inapplicability of test norms: Critics of TIERS have noted that the use of test norms that are developed cross-sectionally may produce a modest positive bias (1 NCE) when compared to longitudinal measures of program gains. In addition, although test norms ideally are based on representative samples of students, in practice the refusal of districts to participate in test norming studies make test norms unrepresentative. According to Linn (1982), "What remains unclear is whether the differences due to nonrepresentativeness are consistent across publishers, and, if so, whether the bias this introduces inflates or deflates the estimates of gains" (p.13).

- Selection biases: The proper implementation of Model A requires that the pretest not be used in any way to make selection decisions. Linn (1982) found anecdotal evidence that this prohibition is sometimes violated. For example, a student may be selected to receive Title I services based on a different test or teacher judgment, but his scores on the subsequent Model A pretest indicated that he does not need Title I services. Teachers or administrators might then make the logical educational decision to remove the student from the program. There are no comprehensive data on the frequency of this practice; however, even if the impact of such practices is small, Linn points out that "it adds to the cumulative bias in the direction of overestimating the impact of Title I" (p.15).

- Nonrandom attrition: Students who leave compensatory education programs prior to the end of the school year usually are not available for posttesting. To the extent that attrition is not random, bias of undetermined direction results. For example, more able students may "graduate" from the program before year's end so that needier students can be served. If posttest scores are not obtained from these presumably higher scoring students, estimates of gains will be underestimated.

In addition, the following practical problems lead to biased results--usually overestimates:

- Conversion errors: As test results are aggregated in the TIERS system, errors can creep in, especially as teachers and administrators manually convert raw scores to percentiles and percentiles to NCEs. Investigation shows that errors are more likely to be in favor of increased (rather than decreased) growth, which can amount to a 1 NCE overestimate, on average.

- **Test administration: Because it is important to teachers and administrators to show positive program results, pretests and posttests may not be administered under the same conditions. Thus, subtle--and sometimes not so subtle--bending of the rules can take place. For example, the teacher might encourage all students to take the pretest but not make the maximum effort to ensure that low scoring students take the posttest. The frequency of such practices is unknown, but their occurrence may add to overestimation problems of TIERS.**

- **Practice effects: A related source of positive bias is the extent to which gains are simply measures of improved test taking skills or successful teaching to the test. While it is important that the test selected for measuring Title I gains should be closely matched with the content and objectives of the project (a poorly matched test would underestimate program performance), Linn (1982) points out "that some inflation of gains is to be expected as the result of learning that is specific to the test" (p.19).**

Linn points out that many of these sources of overestimate are cumulative and "probably account for most of the relatively large estimated gains obtained for the bulk of the 1979-80 results which were based on a fall-to-spring testing pattern" (p.20). Although fall-to-spring testing was the predominant pattern in the early TIERS data, the 1983-84 data presented in this report are based on annual tests. While annual data still are subject to overestimates resulting from, for example, regression effects, the estimates of Title I gains based on this testing cycle are more likely to be a true picture of the impact of Title I.

TIERS data have other weaknesses, although they would not have a systematic effect on estimates of program outcomes. First, because participation in the TIERS system is voluntary, data are not available from all States. In 1984-85, 41 of 53 SEAs reported Statewide data in a format that could be aggregated at the national level (Reisner and Marks, forthcoming). Second, because States are allowed to sample districts, data often come from a subset of districts in a State. The small number of districts or questionable sampling methods in some States could result in nonrepresentative samples. Finally, the quality of data submitted by some States have been poor. Nevertheless, the achievement patterns reported by TIERS are consistent

from year to year suggesting that these problems of data quality do not systematically bias estimates of achievement upward or downward.

The National Assessment of Educational Progress

I. Background and Study Characteristics

A. Overview of the Study

The NAEP--which was originally sponsored by the Carnegie Corporation--is a Federally funded effort to assess national achievement in reading, mathematics, science, social studies, and other learning areas. The central goal of NAEP is to assess periodically the educational achievement of the nation's youth and changes in achievement. Begun in 1969, NAEP periodically collects cross-sectional data on the achievement of 9, 13, and 17-year-olds and young adults between the ages of 25 and 35.

B. Sampling Procedures

NAEP data are derived from a multi-stage sample. The Primary Sampling Units (PSUs) have been stratified by region, State, community size, and socioeconomic status for the smallest size categories. Schools are randomly sampled within PSUs, and 10 to 35 students are randomly sampled from selected schools.

NAEP sample sizes are large. For example, roughly 29,000 9-year-olds, 41,000 13-year-olds, and 36,000 17-year-olds were involved in the 1978 assessment of reading.

C. Student, Program, and Outcome Measures

Since the major goal of NAEP has been to assess educational progress, the major data collection effort has been focused on achievement outcomes. Outcome data have been collected in 10 areas: art; career and occupational development; citizenship; literature; mathematics; music; reading; science; social studies; and writing. As of 1983, all of these areas had been assessed at least once. Citizenship, social studies, art, and

music had been assessed twice. Reading, writing, mathematics, and science had been assessed three times.

Learning objectives are developed and revised over time in consultation with members of the educational community. Test items and exercises are written to assess these objectives. Tests and other measurement instruments are administered by NAEP staff (the contract for conducting NAEP was previously held by the Education Commission of the States and is currently held by the Educational Testing Service) to minimize burden on school district and provide for as nearly standardized test administration procedures as possible. Although new test items and exercises are added for each reassessment, common items are retained to permit analysis of changes in achievement over time.

NAEP collects limited data on student background variables. The following scaled background variables have been collected: parent education, hours spent on homework, parent's occupation, presence or absence of items in the home, gender, and race/ethnicity. No information is available through NAEP on program variables.

II. Strengths and Weaknesses

One major strength of the NAEP is the quality of the achievement data collected. Care is exercised to identify learning objectives, match test items to these objectives, and administer assessment instruments in a consistent manner. At the same time, since test items are not and cannot be matched to specific curricula and teaching methods in the nation's schools, underestimates of achievement levels and changes may result.

Another strength is the sampling design for this data base. However, it is important to note that some groups are systematically excluded, which probably results in overestimates of achievement. First, only students in school are eligible for selection. Thus the 17-year-old sample excludes dropouts. In addition, handicapped students and those with limited English proficiency are excluded. Finally, school

districts can, at their discretion, exclude students who cannot be properly assessed. These limitations in the NAEP samples have implication for the usefulness of these data for evaluating Chapter 1 since the excluded groups overlap with the population targeted to receive Chapter 1 services.

Another shortcoming of the NAEP data for assessing Chapter 1 is that they do not indicate which students receive compensatory education services or what the characteristics of such services are.

Finally, as Koretz (1986) points out, NAEP data are not particularly useful for analyzing trends in test score data. One reason for this problem is that there has been no formal equating of scores from one assessment to the next. In addition, NAEP has traditionally reported results in terms of percents of students correctly answering an item. Finally, NAEP usually does not report standard deviations.

The CBO Report on Trends in Educational Achievement

I. Background and Study Characteristics

A. Overview of the Study

Trends in Educational Achievement is a report prepared by the Congressional Budget Office (CBO) at the request of Congress. (Daniel Koretz of the CBO had lead responsibility for the report.) The report documents and investigates the decline and subsequent rise in trends of national educational achievement. The study is based on reanalyses of existing data from the Scholastic Achievement Test (SAT), the American College Testing Program (ACT), NAEP, the National Longitudinal Study of the High School Class of 1972 (NLS), the High School and Beyond study (HSB), and annual test data from the State of Iowa.

The report discusses: topics in aggregate trends (e.g., when the decline began, how long the decline has lasted, what the depth of the decline was, and when the

decline was reversed) and differential trends among groups of the school-age population (e.g., differences in age group trends, in trends for boys and girls, in trends for high and low achievers, and in trends for minority groups and whites).

B. Sampling Procedures

Since data for this report came from existing sources, sample characteristics must be linked to the original sources--either testing programs or studies. Some of these sources derived data from nationally representative or nearly representative samples. (For example, see the discussion of the NAEP sample, discussed previously.) Other data sources clearly have nonrepresentative samples. For example, the SAT and the ACT testing programs provide data on subsets of the school-aged population that seek college entrance. The Iowa Statewide test data--although representative of that State's students--are not nationally representative.

C. Student, Program, and Outcome Measures

Since this is a study of achievement trends, data sources were selected that included student achievement data over time. In some cases, such as the SAT data and the Iowa State data, trend information is available over decades. Other sources--such as the NAEP--have only two or three measurement points, which make them less useful for trend analysis. Although all data sources for the study provide some measure of students' educational achievement, the details of what is measured vary because the data sources are employed for different purposes. The SAT, for example, is aimed at predicting college success, and its developers see it as less appropriate as an achievement measure than the ACT.

Data sources vary in the amount of student background data collected. Some sources (such as the NLS and HSB) have fairly extensive background data; others such as the SAT and the Iowa data have little or no information on test takers.

Program data from these sources are limited to distinctions, for example, between students who attend public or private schools or who attend schools in disadvantaged communities.

II. Strengths and Weaknesses

The report is an excellent summary of national achievement trends; but, because it relies solely on existing data sources, its strengths and weaknesses depend in the end on the strengths and weaknesses of those sources. The report acknowledges those strengths and weaknesses and is circumspect about findings and conclusions on less than optimal data. The following summarizes acknowledged weaknesses in the main data sources for the report (for strengths and weaknesses of the NAEP, see the discussion above):

- The SAT: In addition to the nonrepresentativeness of the sample and the fact that the SAT is primarily designed to measure achievement, a problem with the use of these data for trend analysis is that, although there are many years of SAT results, there are inconsistencies in the results over time. For example, tabulations for recent years are based on the most recent test that a given student took. Tabulations of results from the mid 1950s to mid 1960s contain all test results of each test taker.

- The ACT: This data source has the advantage over the SAT that it is intended to measure achievement as well as predict college success. Like the SAT, the ACT is somewhat inconsistent over time, which is a disadvantage for trend analysis.

- The NLS and HSB: Both of these studies provide extensive background data on students. At the same time, the utility of both the NLS and HSB is limited for trend analysis because only a few measurement points are available. Evidence from other sources suggests that achievement trends were nonlinear during the timeframes of these studies; thus use of the NLS or HSB to examine these trends probably leads to analytic distortions.

- The Iowa Statewide Data: The major strength of these data for examining achievement trends is that they provide three decades of annual equated test results. However, their usefulness is limited because, as already mentioned, the Iowa student population is not representative of students nationwide. In addition, the

Iowa data do not provide sufficient information on student backgrounds to examine differential achievement trends across subgroups of students.

The Sustaining Achievement Study

I. Background and Study Characteristics

A. Overview of the Study

The SAS was conducted by staff of the Chapter 1 Technical Assistance Centers, which are funded by ED to provide assistance to states and LEAs in conducting Chapter 1 evaluations and other technical activities. The overriding purpose of the study was to provide State and local evaluators with information on how to implement evaluations of the sustained effects of Chapter 1, as the Chapter 1 statute requires. The study used existing data to address the following questions:

- Are there different patterns of achievement for students with varying patterns of participation in compensatory education?
- Do achievement patterns vary across grade levels, between reading and mathematics programs, and for subtest scores or specific test items?
- Do achievement patterns differ depending upon whether results are based on fall-spring or spring-spring testing schedules?

The data for the SAS came from three sources: (1) a review of locally conducted sustained effects evaluations, (2) a reanalysis of test results from 17 school districts, and (3) reanalysis of data from the Sustaining Effects Study, with particular attention to trends in individual test items. The review of the SAS will focus on the second data source and its reanalysis.

B. Sampling Procedures

Data for the reanalysis of local test data came from 17 school districts that had the required data available and were willing to participate. TAC staff collected and analyzed spring 1982 and spring 1983 data for approximately 66,500 students in

Chapter 1 schools. The authors acknowledge that although "these systems were selected primarily for the availability of their information, and they did not constitute a representative national sample, the overall achievement levels and participation percentages in Chapter 1 for students in the sample appeared quite similar to overall national figures" (Gabriel et al., 1985, p. 2)

C. Student, Program, and Outcome Measures

The TAC staff limited test data to students in grades 2 through 6 since students in these grades make up roughly 70 percent of all Chapter 1 students. Test data were also limited to the two most used instruments in Chapter 1 programs: the California Achievement test and the Science Research Associates test.

II. Strengths and Weaknesses

The SAS is an example of the useful work that can be done with a limited budget using existing data. The strengths of the study are its documentation of the biases resulting from a fall-spring testing cycle. (See the discussion of this problem in the foregoing description of the TIERS.) In addition the study identifies evidence of consistent longitudinal achievement patterns for four groups of students: those who never enter Chapter 1 ("out/out"), those who received Chapter 1 services one of the two years under consideration ("in/out" and "out/in"), and those who receive services both years ("in/in").

At the same time, the SAS--like any reanalysis of existing data --suffers from the weaknesses of those data, which the authors acknowledge. In this instance, these problems result from the necessity of depending on districts to cooperate and to supply the data they have. As already mentioned, this prohibits any generalization to Chapter 1 programs as a whole. Thus, although the SAS has produced findings that provide a useful perspective on the sustained effects of Chapter 1, the findings are only speculative and require confirmation from nationally representative data. A related

problem is that the data came from only 17 cooperative districts that had data available, and over 75 percent of these data came from just four districts. If one or two of these districts were unusual or had particularly flawed data, the generalizability and quality of the data could be compromised even further. Because the TAC staff had to depend on locally collected data, they had no control over the quality of testing and other data collection procedures. Although they may have been able to control for some of these problems by limiting the number of tests considered, variability in local practices can have undetermined influences on evaluation results. (See the discussion of TIERS for more detail on possible problems.)

The need to examine longitudinal results also led to problems for the SAS. First, the study team was able to examine only two data points. Although their results are consistent across districts and across grade levels, even two or three more data points would help confirm and elaborate their trends. Second, as the authors admit, the requirement of two test scores for each student limits the result to the subsample of students who were present for both tests. Given the limits of the data, no estimate of the bias from student attrition is possible.

Summer Learning and the Effects of Schooling

I. Background and Study Characteristics

A. Overview of the Study

Undertaken in the early 1970s, Summer Learning and the Effects of Schooling (Heyns, 1978) was a study aimed at disentangling the influences of socioeconomic status (SES) and schooling on achievement. The basic design of the study used summer vacation as a "natural experiment" to help determine the impact of social class when comparatively little organized schooling occurs. Heyns constructed a longitudinal data base from retrospective test scores for approximately 3,000 sixth and seventh graders

attending the Atlanta public schools. In addition, she collected questionnaire data from parents and information about summer experience. Her major finding contradicts much of the 1960s' and 1970s' literature on the effects of schooling: "Schooling makes a substantial contribution to cognitive growth....Schooling apparently attenuates the influence of socioeconomic status on achievement and thereby reduces the direct dependence of outcomes on family background" (p.187).

B. Sampling Procedure

The data for the study were collected from a stratified random sample of sixth and seventh graders in Atlanta. Schools were stratified by race and by proportion of free lunch recipients (a measure of socioeconomic status). Within each stratum, schools were ranked by enrollment within each stratum and systematically selected from the strata. This resulted in a sample of 42 schools and a total sample size of approximately 4,800 students. The stratifying strategy resulted in a sample that was racially balanced and diverse regarding the SES of both blacks and whites. For further assurance of variability in SES, schools with high proportions of low income whites or with high income blacks were oversampled.

The original sample was reduced by less than perfect response rates to the parent questionnaire and by students who did not have test scores for one or more of the three testing periods. The final sample of students--those with a valid parent questionnaire and three consecutive test scores--represented approximately 60 percent of the original sample (2,978 students). An additional 739 seventh graders had four consecutive test scores. Heyns chose to analyze these data separately in several instances.

C. Background, Program, and Outcome Measures

Interviews with parents provided information on student background data. The response rate for the interviews was 86 percent. These interviews obtained information

on standard indicators of socioeconomic status such as mother's and father's education, occupational status, income level, and home ownership. These interviews were also the main source of program information--in this case, information about summer school and other summer activities.

The intermediate battery of the Metropolitan Achievement Test (MAT) was the outcome measure for this study. Because this was the districtwide achievement test for the Atlanta school system, longitudinal outcomes could be obtained retrospectively and analyzed. Although students in the sample were sixth or seventh graders when the study was done, test scores for the sixth graders actually came from their fifth-grade fall and spring test dates and from the their sixth-grade fall test date (fall of 1971 and spring and fall of 1972). Seventh graders' scores came from their fifth-grade spring test date (spring 1971) and from their sixth-grade fall and spring tests (fall 1971 and spring 1972). For both cohorts, the summer period used for analysis was between fifth and sixth grade (summer of 1971 and summer of 1972).

II. Strengths and Weaknesses

Overall, the quality of this study is very high. The study design and execution rest on a firm conceptual base, which is clearly articulated in the study. The sample design achieves variability on background variables of interest. One possible problem with the sample is the relatively high attrition rate of 40 percent. However, the author shows that attrition appears to be random--at least with respect to race. (The final sample was made up of 60.6 percent of the whites in the original pool and 60.7 percent of the blacks.)

Care was taken in collecting student background data from parents. Just the fact that family income data and other SES indicators were collected from parents sets this study apart from other attempts to assess the effects of schooling. Further care was taken to attempt to interview in person parents who had no telephone, thus avoiding

an obvious bias in favor of students from higher SES backgrounds. Two other strengths of the study were its use of longitudinal data and its analysis of data at the student level. Studies of school effectiveness often have relied on cross-sectional analysis and aggregated data.

A final strength is that the study used extant data imaginatively to address several important policy issues. The longitudinal data base permitted creation of a natural control situation--summer vacation--to test the independent impact of schooling on cognitive growth. Employing test scores from a battery of tests, Heyns was able to examine the impact of schooling on a number of subject areas. In addition, by collecting data on summer activities, she was able to explore in depth the nature of summer learning.

APPENDIX E

Measures Used in This Report

Over one-half dozen measures have been used in this report to describe how well students did on achievement tests. Vertical scale scores, NCEs, percentiles, reading proficiency scores, and more have been cited to report whether one group of students learned more than another or how much a group of students has learned. All of these types of achievement scores are only different ways of expressing a simple fact, that is, how many test items were answered correctly compared to some standard.

Percentiles

Percentiles are the most commonly used scores for expressing how well a student or group of students has performed compared to other students. A percentile (sometimes called a percentile rank) is a score which expresses how well a given student or group of students has done compared to other students. For example, if a child's percentile is 40, we say the child did better than 40 percent of all other students on the test; that is, 40 percent of the students taking that test got fewer of the items correct. But percentiles have some limitations, the most significant of which is that a percentile scale's intervals between points are not of equal size; that is, a percentile scale is not an equal interval scale. For example, it generally takes more of an achievement gain to improve from the 20th percentile to the 30th than it takes to improve from the 40th to the 50th simply because of the properties of the scale. Further, because the percentile scale is not an equal interval scale, it is not correct to use percentiles in many types of arithmetic operations, such as is necessary to calculate averages.

Normal Curve Equivalents

Since averages across students' scores are generally needed for program evaluations, an alternative to percentiles called the Normal Curve Equivalent (NCE) was developed. With NCEs, the students' scores could be averaged (permitting large scale

evaluation), they could be converted easily from percentiles to NCEs and back (permitting simplified reporting), and they used a 100-point scale with an average of 50 (permitting an easier grasp on the "meaning" of the scale). The NCE scale was subsequently mandated for use in TIERS and is still in use under Chapter 1.

As with percentiles, it is assumed that NCE scores will remain the same for a student or group of students in the absence of specific interventions. That is, all students will learn enough in the normal course of events to maintain their positions relative to one another, which means they are expected to get the same NCE score each subsequent time they take a test. An NCE gain or loss reflects a change in position relative to other students.

Standard Scores

Numerous other scales exist or can be created to express how well students did on a test compared to other students which assume that students' relative positions will remain constant despite the fact that learning has occurred. (It should be noted that no assumption of "equal" learning is involved; in fact, a child scoring at a higher percentile rank or NCE score is probably going to learn more in an absolute sense between tests than a child who originally scored lower; all that counts with these measures is relative position.) These measures fall within a family which is often given the general label of standard scores.

All standard scores, of which NCEs are merely one example, are based on known qualities of the normal ("bell-shaped") curve. Different scores in this family result from multiplying the standard deviation of the raw scores by different constants and adding different constants to the mean of the raw scores. Standard deviations measure the extent to which the scores of different children are dispersed around the mean or average score. If the scores are distributed normally, i.e., they describe a bell-shaped curve, then by definition a little more than 68 percent of the scores will fall within

one standard deviation above or below the mean, another 27 percent will fall within the second standard deviation, and a little less than 5 percent of the scores will fall further out from the mean than the second standard deviation.

Based on the same raw score and depending on which mean and standard deviation "transformations" are used, one can report scores in terms of standard deviation units (the number of standard deviations, e.g., 1.25, above or below the mean), Z scores (with a mean of zero and a standard deviation of 1), t scores with a mean of 50 and a standard deviation of 10), or NCE scores (mean of 50, standard deviation of 21.06). The different scores are only the result of different transformations; with a little arithmetic, any member of this family can be converted into another.

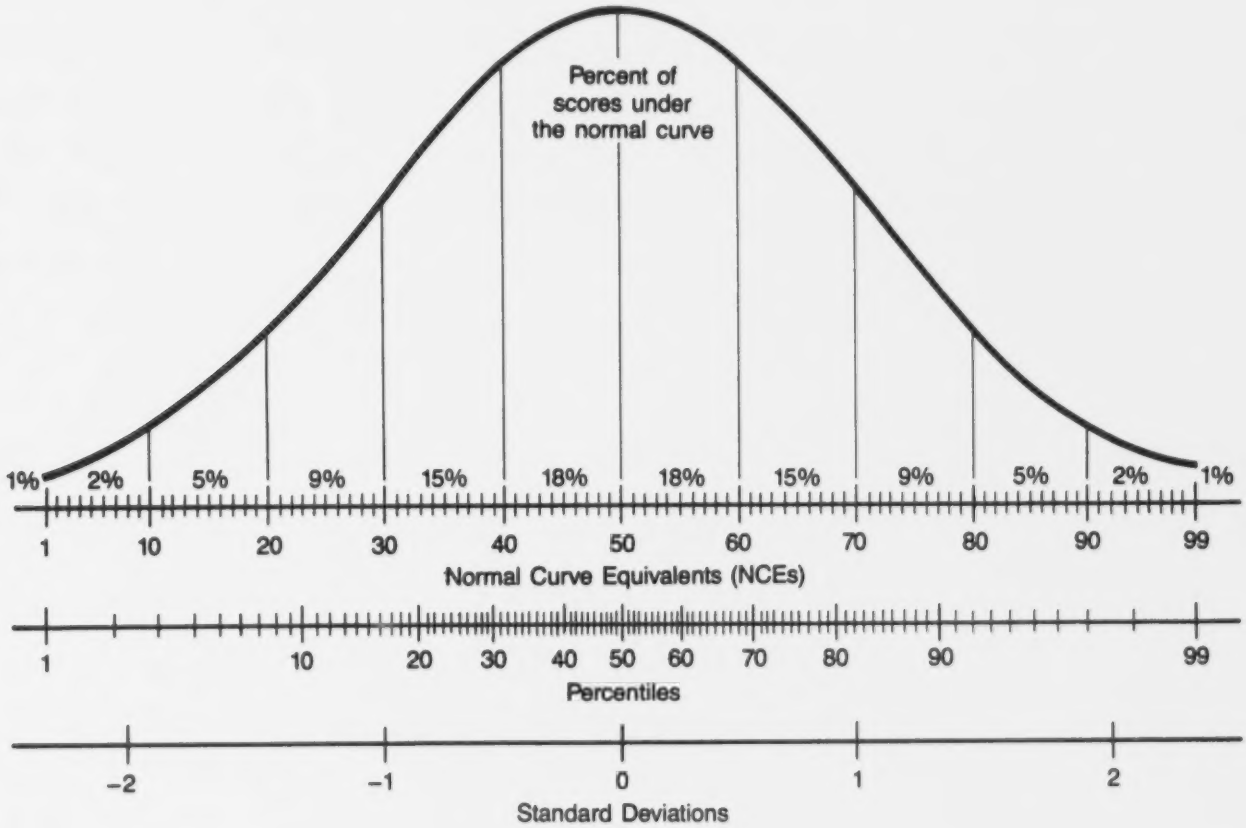
Figure E.1 illustrates the relationships among percentiles, NCEs, standard deviation units, and the normal curve. The figure shows, for example, that an NCE of 29 corresponds to a percentile rank of about 16 and lies about one standard deviation below the standardized mean of 0 (NCE mean of 50). Standard scores provide a means of reporting the relative position of test-takers; they say nothing directly about how much more one student or group of students knows than another. To report those findings requires a member of the family of expanded scale scores.

Expanded Scale Scores

Students are expected to gain knowledge between one test and the next just as older students, all other things being equal, are expected to know more than younger ones. Expanded scale scores have been developed to allow reporting how much more one student or group of students knows than another or how much a student or group has learned from one test to the next.

Figure E.1

Comparison of Test Scores



Reading Proficiency Scores

NAEP's reading proficiency score, the vertical scale score used in the Sustaining Effects Study, and the expanded standard scores reported by NIE are examples used in this report which are members of the expanded scale score family.

The expanded scales are developed generally as part of test-norming procedures and often involve administering some test items to students in grades above and below the grade level for which the items were designed. This permits subsequent estimations of the likelihood of marking the items correctly. These likelihoods can, in turn, be used to construct a common linear scale which cuts across test levels and which is composed of scores which approximate an equal interval scale. Students' performances on the test, regardless of the level of the test they took, can then be placed on this scale to determine how much they know. As the children learn more, their position on the scale gets higher, reflecting their greater knowledge.

If the expanded scale can be linked to some agreed-upon standard, such as the reading skill levels used by NAEP to describe the practical meaning of their reading proficiency scores (e.g., rudimentary, basic, intermediate, adept, advanced), then knowing how much further up the scale a student or group of students has moved provides extremely useful information on program success. If, however, no external linkage to an agreed-upon standard is available, then judging the amount of achievement growth also requires looking at comparisons with other students.

Grade Equivalent Scores

Grade equivalent scores had been the most common measures for reporting performance on standardized tests until the last few years. Grade equivalents assume that levels of knowledge can be described in terms of equivalent amounts of time in school. For example, a fifth grader at the start of the school year doing "average" work would be expected to obtain an "on grade level" score of 5.0, while a student in

the same class who might not be doing as well may score "below level" at, say, 4.2. Grade equivalent scales are not equal interval scales so they cannot be used with confidence to report how much more one student or group knows than another. Further, despite the logical sound of a statement such as, "The third graders were reading one and one-half years below grade level," the sense disappears on closer examination. In this case, for instance, it is very unlikely that students midway through the first grade would be able to take the test at all. In fact, the farther scores are from average (and students in Chapter 1 often score well below average), the greater the amount of distortion in meaning found in grade equivalents.

Summary

Different test scores are designed to accomplish different purposes, and generally speaking, test scores fall into one of two families. The first family is based directly on the properties of the normal distribution and is used to compare the relative positions of test takers. Examples of members of this family used in this report are percentiles and normal curve equivalents (NCEs). The second family is based on relationships across levels of a test which permit developing equal interval scales. With an equal interval scale tying together the different levels of a test, it becomes possible to report how much students have learned from one time to the next or how much more one group of students knows than another. Among members of this family cited in this report are reading proficiency scores and expanded standard scores.

APPENDIX F

One-Year Achievement Gains Expressed in Normal Curve Equivalents Rather than Percentile Ranks

This appendix presents the information contained in Figures 3.2, 3.3, 3.4, and 4.4 expressed in terms of Normal Curve Equivalents (NCEs) rather than in terms of percentile ranks as they appear in the text.

Percentile ranks indicate the percentage of students nationwide that obtained scores lower than that particular achievement level.[1] A percentile rank of 50, for example, indicates the achievement level which 50 percent of all students scores below. Percentile ranks must be interpreted cautiously because they are not based on an equal-interval scale. This means that a child who increases achievement by 10 percentile points, from 45 to 55 for example, has not learned the same amount as a child who increases from 85 to 95. Percentile ranks cannot therefore be used for averaging student achievement gains since each point of gain does not represent the same amount of learning.

Another measure which is derived from percentile rankings, the Normal Curve Equivalent (NCE), is an equal-interval scale. As a result, NCEs legitimately can be summed and averaged to compare the gains of different students or groups of students.

All of the percentile ranks in Figures 3.2, 3.3, 3.4, and 4.4 in the body of the report were based on data calculated using NCEs or other standard scores. These scores were converted to percentile ranks for purposes of presentation.

---

[1]The Chapter 1 Technical Assistance Centers (TACs) have materials that explain different types of measurement in terms suitable for lay audiences.

Figure 3.2

Gains in NCE Scores for Chapter I Students in Reading and Mathematics, 1983-84

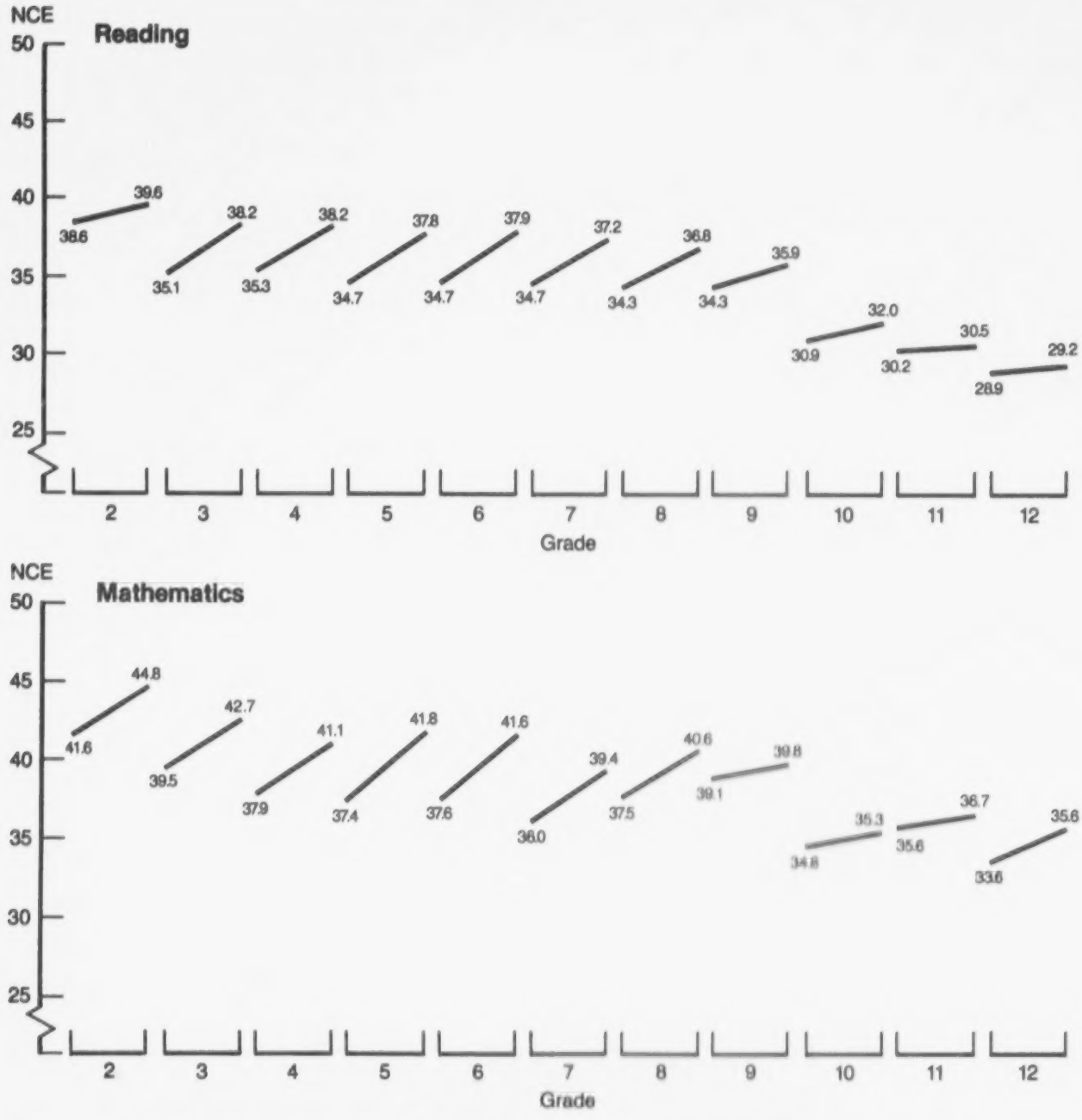


Figure reads: From spring 1983 to spring 1984, the NCE scores of second grade students who received Chapter I reading instruction increased from 38.6 to 39.6, while the NCE scores of 12th grade students remained fairly constant around 29.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology. 1985.

Figure 3.3

Gains in NCE Scores for Title I Students in Reading and Mathematics, The Sustaining Effects Study, 1976-77*

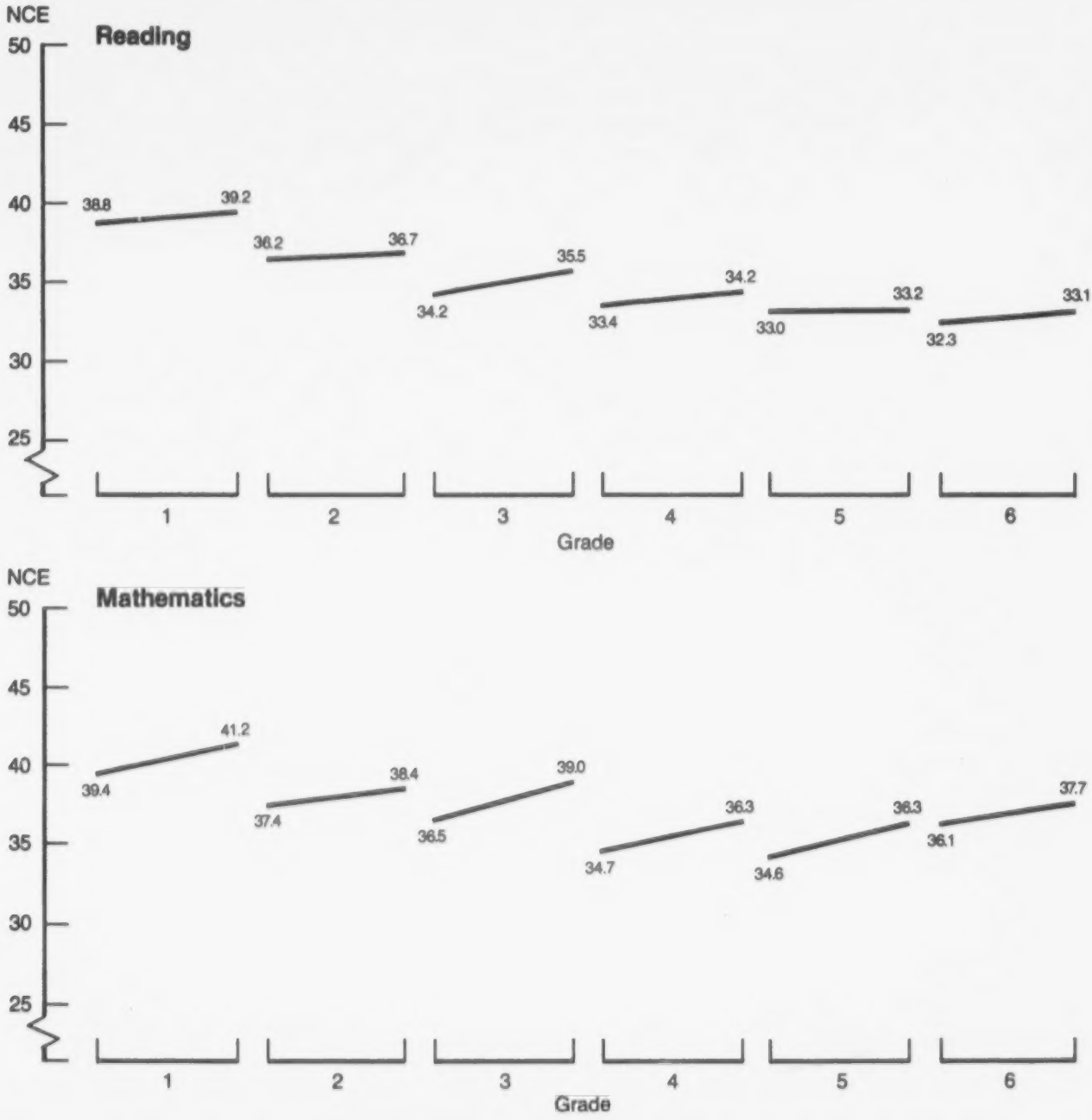


Figure reads: From the fall to the spring testing, 1st grade students enrolled in Chapter I reading instruction moved from just below an NCE score of 39 to just above it.

*NCE scores presented are based on scores from a fall-spring testing cycle in contrast with the spring-spring cycle used for TIERS data in Figure 3.2.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, *Report 10: Compensatory Services and Educational Development in the School Year.* Santa Monica, CA: System Development Corp., 1981.

Figure 3.4

Gains in NCE Scores for Title I Students and Similar Students Not Receiving Compensatory Education, The Sustaining Effects Study, 1976-77*

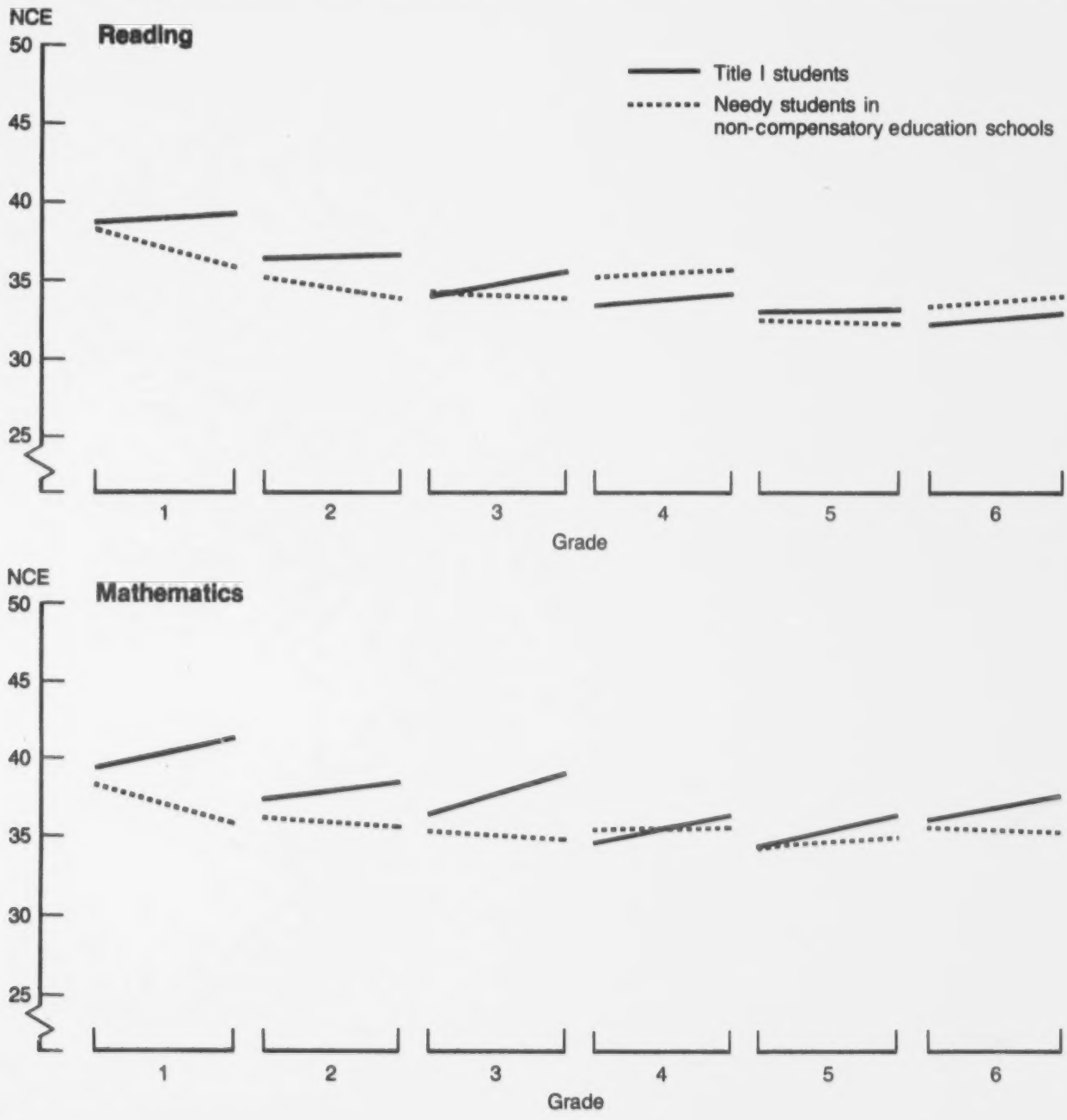


Figure reads: From the fall to the spring testing, 1st grade Title I students receiving reading instruction moved from just below an NCE score of 39 to just above, while needy students in non-Title I schools dropped from just above an NCE score of 38 to just below a score of 36.

*NCE scores presented are based on scores from a fall-spring testing cycle in contrast with the spring-spring cycle used for TIERS data in Figure 3.2.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, *Report 10: Compensatory Services and Educational Development in the School Year.* Santa Monica, CA: System Development Corp., 1981.

Figure 4.4

Comparison of Gains in NCE Scores for Reading and Mathematics by Testing Cycle, 1983-84

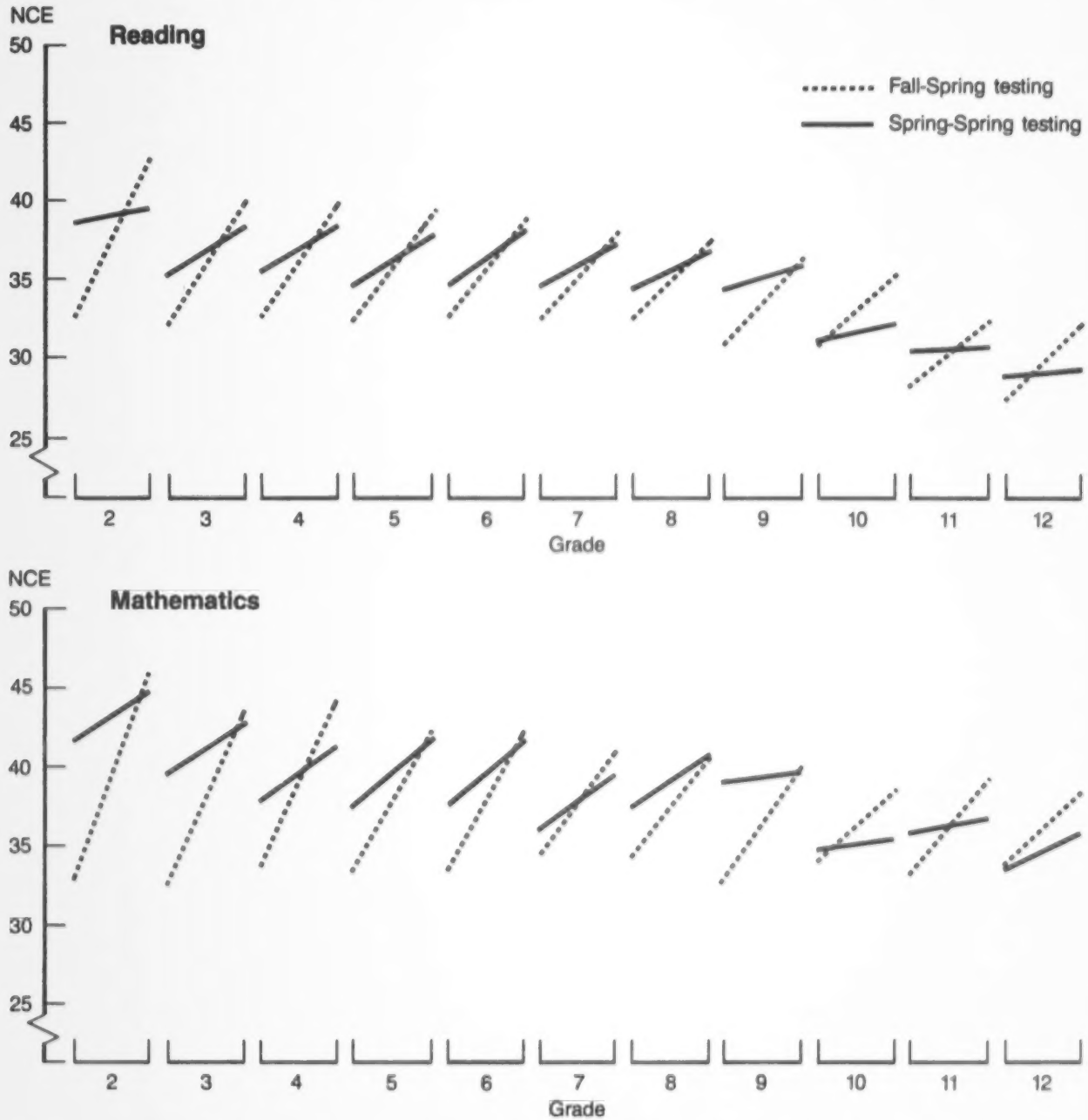


Figure reads: The NCE scores of the average second grader receiving Chapter I reading instruction increased from 32.7 to 42.6 using a fall to spring testing cycle, while a spring to spring testing cycle resulted in a less dramatic NCE score shift from 38.6 to 39.6.

Source: M. A. Carpenter and P. A. Hopper, *Synthesis of Chapter I Data: Summary Report.* Reston, VA: Advanced Technology, 1985.

TABLE F.1

Normal Curve Equivalents (NCEs) and Percentile Ranks for Figures 3.2 and 4.4

1983-84 Chapter 1 Reading and Mathematics Achievement Results

	Full Year				Fall-Spring			
	NCEs		Percentiles		NCEs		Percentiles	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Reading								
Grade 2	38.6	39.6	29	31	32.7	42.6	21	36
3	35.1	38.2	24	29	31.9	40.0	20	32
4	35.3	38.2	24	29	32.4	39.8	20	32
5	34.7	37.8	23	28	32.4	39.0	20	30
6	34.7	37.9	23	28	32.5	38.8	20	30
7	34.7	37.2	23	27	32.3	37.9	20	28
8	34.3	36.8	23	27	32.4	37.5	20	28
9	34.3	35.9	23	25	30.8	36.1	18	26
10	30.9	32.0	18	20	30.6	35.1	18	24
11	30.2	30.5	17	18	28.1	32.2	15	20
12	28.9	29.2	16	16	27.2	31.9	14	20
Mathematics								
Grade 2	32.8	45.7	21	42	32.8	45.7	21	42
3	32.5	43.3	20	38	32.5	43.3	20	38
4	33.8	44.1	22	39	33.8	44.1	22	39
5	33.5	42.2	22	36	33.5	42.2	22	36
6	33.8	42.2	22	36	33.8	42.2	22	36
7	34.7	41.0	23	34	34.7	41.0	23	34
8	34.3	40.3	23	32	34.3	40.3	23	32
9	32.8	40.0	21	32	32.8	40.0	21	32
10	34.1	38.3	23	29	34.1	38.3	23	29
11	33.2	39.1	21	30	33.2	39.1	21	30
12	33.9	38.4	22	29	33.9	38.4	22	29

Table reads: The percentile ranks corresponding to the second grade reading lines presented in Figures 3.2 and 4.4 are 29 and 31 for the pretest and posttest, respectively. The corresponding NCE scores are 38.6 and 39.6.

Source: M. Carpenter and P. Hopper, Synthesis of Chapter 1 Data: Summary Report. Reston, VA: Advanced Technology, 1985.

TABLE F.2

Normal Curve Equivalents (NCEs) and Percentile Ranks for Figures 3.3 and 3.4

Fall-to-Spring Reading and Mathematics Achievement Scores for Title I Students and Similar Students Not Receiving Compensatory Education

		NCEs		Percentile Ranks	
		Pretest	Posttest	Pretest	Posttest
Reading					
Grade 1	Title I Students	38.8	39.2	29.7	30.4
	Needy Students	38.2	35.9	28.8	25.2
2	Title I Students	36.2	36.7	25.6	26.4
	Needy Students	35.2	34.0	24.1	22.4
3	Title I Students	34.2	35.5	22.6	24.6
	Needy Students	34.5	34.0	23.1	22.4
4	Title I Students	33.4	34.2	21.5	22.7
	Needy Students	35.1	35.6	23.9	24.7
5	Title I Students	33.0	33.2	21.0	21.2
	Needy Students	32.6	32.4	20.4	20.2
6	Title I Students	32.3	33.1	20.1	21.1
	Needy Students	33.7	34.0	21.9	22.4
Mathematics					
Grade 1	Title I Students	39.4	41.2	30.8	33.8
	Needy Students	38.3	35.9	29.0	25.1
2	Title I Students	37.4	38.4	27.5	29.1
	Needy Students	36.2	35.5	25.6	24.5
3	Title I Students	36.5	39.0	26.1	30.1
	Needy Students	35.4	34.9	24.4	23.7
4	Title I Students	34.7	36.3	23.4	25.7
	Needy Students	35.5	35.8	24.5	25.0
5	Title I Students	34.6	36.3	23.2	25.8
	Needy Students	34.6	35.0	23.3	23.8
6	Title I Students	36.1	37.7	25.5	28.0
	Needy Students	35.5	35.1	24.6	24.0

Table reads: The percentile ranks corresponding to first grade Title I reading students' lines in Figures 3.3 and 3.4 are 29.7 and 30.4 for the pretest and posttest, respectively. The corresponding NCE scores are 38.8 and 39.2.

Source: M. Wang, M. Bear, J. Conklin, R. Hoepfner, Report 10: Compensatory Services and Educational Development in the School Year. Santa Monica, CA: System Development Corp., 1981.

APPENDIX G

Sample Sizes for Figures and Tables

TABLE G.1

Sample Sizes for Figures 3.1, 3.2, 3.3, 3.4, 4.2 and 4.3 and Table 4.1

Grade/Subject		Figure 3.1/ Figure 4.3 (Spring-Spring Testing	Figure 3.2/ Figure 3.3 Figure 3.4 (Title I Students in Title I Schools)	Figure 3.3/ Figure 3.4 (Need Compensatory Education in Non-Title I Schools)	Figure 3.4/ (Non-Compensatory Students in Title I Schools)	Table 4.1/ (Compensatory Education Students)	Table 4.1/ (Non-Compensatory Education Students)	Figure 4.2/ (Attended Summer School)*	Figure 4.2/ (Not Attended Summer School)*	Figure 4.3/ (Fall-Spring Testing)
1	Reading		2,785	541	2,115	344	296			
	Math		1,598	1,051	2,468	97	435			
2	Reading	93,959	3,036	678	1,754			324	1,013	182,490
	Math	54,790	1,686	1,035	2,207			201	544	63,922
3	Reading	115,160	3,022	591	1,761	512	305	371	1,136	158,221
	Math	64,629	1,793	1,057	2,126	306	178	229	560	68,215
4	Reading	119,437	2,392	613	1,876			498	1,132	140,961
	Math	72,558	1,434	1,169	2,178			345	664	68,328
5	Reading	121,383	2,227	548	1,992			315	972	121,558
	Math	77,677	1,331	1,270	2,226			241	593	63,350
6	Reading	105,021	1,982	676	1,992			235	831	106,666
	Math	68,235	1,134	1,231	2,272			173	578	55,456
7	Reading	65,246								69,429
	Math	39,072								36,483
8	Reading	65,826								49,866
	Math	45,842								28,589
9	Reading	31,349								30,818
	Math	22,635								18,012
10	Reading	13,489								17,992
	Math	8,372								7,485
11	Reading	7,967								9,737
	Math	5,096								3,297
12	Reading	4,506								5,873
	Math	3,352								1,859

Table reads: The grade 2 reading achievement values reported in Figures 3.1 and 4.3 for Spring to Spring testing are based on 93,959 students.

* Grade at last test period.

1/ M. A. Carpenter and P. A. Hopper, Synthesis of Chapter 1 Data: Summary Report. Reston, VA: Advanced Technology, 1985.

2/ M. Wang, M. Bear, J. Conklin, R. Hoepfner, Report 10: Compensatory Services and Educational Development in the School Year. Santa Monica, CA: System Development Corporation, 1981.

3/ National Institute of Education, Compensatory Education Study. Final Report to Congress from the National Institute of Education. Washington, DC: September 1978.

4/ L. Klibanoff and S. Haggart. Report 8: Summer Growth and the Effectiveness of Summer School. Santa Monica, CA: System Development Corporation, 1981.

TABLE G.2

Sample Sizes for Achievement of Students With
Different Patterns of Participation in Title I
Across Three Years, Figures 4.4 & 4.5

Participation Pattern			Figure 4.4	Figure 4.5
Year 1	Year 2	Year 3	Reading	Math
Participant	Participant	Participant	1407	685
Participant	Participant	Non-Participant	680	340
Participant	Non-Participant	Participant	358	183
Non-Participant	Participant	Participant	703	321
Participant	Non-Participant	Non-Participant	655	412
Non-Participant	Participant	Non-Participant	520	399
Non-Participant	Non-Participant	Participant	612	384

Table reads: The results presented in Figure 4.4 for three year participants, that is, those who participated in Title I in Years 1, 2, and 3, were based on a sample of 1407 students.

☆U.S. GOVERNMENT PRINTING OFFICE: 1987-174-720/60158